```
                        GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:          January 28, 2004, 20:49:00 ; Search time 12457 Seconds
                                            (without alignments)
                                            11389.146 Million cell updates/sec

Title:           US-10-056-884A-1
Perfect score:   3468
Sequence:        1 caagcactgtgctaaagtgt.........aaaaaaaaaaaaaaaaaaaaa 3468

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
```

```
28:  em_un:*
29:  em_vi:*
30:  em_htg_hum:*
31:  em_htg_inv:*
32:  em_htg_other:*
33:  em_htg_mus:*
34:  em_htg_pln:*
35:  em_htg_rod:*
36:  em_htg_mam:*
37:  em_htg_vrt:*
38:  em_sy:*
39:  em_htgo_hum:*
40:  em_htgo_mus:*
41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                  SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3198.4	92.2	5646	9	AB037738	AB037738 Homo sapi
c	2	2045.4	59.0	182638	2	AC019335	AC019335 Homo sapi
c	3	2043.4	58.9	184589	9	AC008716	AC008716 Homo sapi
c	4	2038.6	58.8	98360	9	AC008473	AC008473 Homo sapi
	5	1640.8	47.3	2412	6	AX405760	AX405760 Sequence
c	6	1161.8	33.5	171949	9	AC008652	AC008652 Homo sapi
	7	1161.8	33.5	209114	9	AC008383	AC008383 Homo sapi
c	8	853.2	24.6	135132	2	AC127249	AC127249 Mus muscu
	9	853.2	24.6	186417	2	AC114984	AC114984 Mus muscu
c	10	819.4	23.6	242679	2	AC117867	AC117867 Rattus no
	11	811.8	23.4	230128	10	AC098707	AC098707 Mus muscu
	12	806.2	23.2	249703	2	AC112599	AC112599 Rattus no
	13	367.4	10.6	781	10	BC049734	BC049734 Mus muscu
	14	319.2	9.2	175059	2	BX323465	BX323465 Danio rer
	15	317.6	9.2	200467	2	BX470157	BX470157 Danio rer
c	16	317.6	9.2	230261	2	BX530085	BX530085 Danio rer
	17	286.4	8.3	184319	2	BX511303	BX511303 Danio rer
	18	284.4	8.2	192400	2	BX530406	BX530406 Danio rer
	19	278.4	8.0	83028	2	BX004755	BX004755 Danio rer
c	20	278.4	8.0	243835	5	AL935304	AL935304 Zebrafish
c	21	255.4	7.4	174712	10	AL831725	AL831725 Mouse DNA
	22	251.6	7.3	1890	5	AY120891	AY120891 Danio rer
	23	220.8	6.4	183038	2	AC107770	AC107770 Mus muscu
c	24	216.4	6.2	246164	2	AC098751	AC098751 Rattus no
	25	216.4	6.2	322972	2	AC129853	AC129853 Rattus no
c	26	212.8	6.1	138872	9	AC131951	AC131951 Homo sapi
	27	189.6	5.5	243299	2	AC133800	AC133800 Rattus no
	28	188	5.4	472	6	BD109391	BD109391 EST and e
	29	167	4.8	3086	9	BC013764	BC013764 Homo sapi
c	30	167	4.8	109201	9	AC000403	AC000403 Genomic s
c	31	167	4.8	169362	2	AL136440	AL136440 Homo sapi
	32	165.4	4.8	251187	2	AL359875	AL359875 Homo sapi
c	33	159.8	4.6	145979	2	AC102815	AC102815 Mus muscu

```
c  34   156.6   4.5 217336   2   AC131344         AC131344 Rattus no
c  35   149.8   4.3    675  10   BC049679          BC049679 Mus muscu
   36   129.2   3.7 256933  10   AC099771          AC099771 Mus Muscu
   37   128.2   3.7 194240   2   AC118142         AC118142 Rattus no
c  38   127.6   3.7 203720   9   AC093861         AC093861 Homo sapi
c  39   114.8   3.3    569   5   AY093634         AY093634 Acipenser
c  40   104.6   3.0 100029  10   AE014174_3        Continuation (4 of
   41    84.4   2.4   2093   5   BC051776         BC051776 Danio rer
   42    81.8   2.4   1450  17   AF132205          Af132205 Homo sapi
   43    81.8   2.4   2689   9   BC032473         BC032473 Homo sapi
   44    80.8   2.3 246548   2   BX530062         BX530062 Danio rer
   45    79.2   2.3   4582   9   BC037864         BC037864 Homo sapi
```

ALIGNMENTS

RESULT 1
AB037738
LOCUS   AB037738   5646 bp   mRNA   linear   PRI 14-MAR-2000
DEFINITION Homo sapiens mRNA for KIAA1317 protein, partial cds.
ACCESSION   AB037738
VERSION   AB037738.1 GI:7243014
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Nagase,T., Kikuno,R., Ishikawa,K.I., Hirosawa,M. and Ohara,O.
  TITLE   Prediction of the coding sequences of unidentified human genes.
            XVI. The complete sequences of 150 new cDNA clones from brain which
            code for large proteins in vitro
  JOURNAL   DNA Res. 7 (1), 65-73 (2000)
  MEDLINE   20181126
   PUBMED   10718198
REFERENCE   2 (bases 1 to 5646)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES   Location/Qualifiers
     source   1. .5646
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="fh13117"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
     gene   1. .5646
                     /gene="KIAA1317"
     CDS   <1071. .2378
                     /gene="KIAA1317"
                     /note="Start codon is not identified."

```
                         /codon_start=1
                         /product="KIAA1317 protein"
                         /protein_id="BAA92555.1"
                         /db_xref="GI:7243015"
                         /translation="QQQKKGTMALSGNCSRYYPREQGSAVPNSFPEVVELNVGGQVYF
                         TRHSTLISIPHSLLWKMFSPKRDTANDLAKDSKGRFFIDRDGFLFRYILDYLRDRQVV
                         LPDHFPEKGRLKREAEYFQLPDLVKLLTPDEIKQSPDEFCHSDFEDASQGSDTRICPP
                         SSLLPADRKWGFITVGYRGSCTLGREGQADAKFRRVPRILVCGRISLAKEVFGETLNE
                         SRDPDRAPERYTSRFYLKFKHLERAFDMLSECGFHMVACNSSVTASFINQYTDDKIWS
                         SYTEYVFYREPSRWSPSHCDCCCKNGKGDKEGESGTSCNDLSTSSCDSQSEASSPQET
                         VICGPVTRQTNIQTLDRPIKKGPVQLIQQSEMRRKSDLLRTLTSGSRESNMSSKKKAV
                         KEKLSIEEELEKCIQDFLKIKIPDRFPERKHPWQSELLRKYHL"
BASE COUNT      1618 a    1169 c    1150 g    1709 t
ORIGIN

     Query Match              92.2%;   Score 3198.4;   DB 9;  Length 5646;
     Best Local Similarity    99.2%;   Pred. No. 0;
     Matches 3225;  Conservative     0;  Mismatches   26;  Indels    1;  Gaps    1;

Qy         183 ACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGA 242
               II  II  IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         760 ACTAAGACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGA 819

Qy         243 AGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCA 302
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         820 AGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCA 879

Qy         303 TCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGAT 362
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         880 TCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGAT 939

Qy         363 TTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTT 422
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         940 TTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTT 999

Qy         423 TCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTA 482
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db        1000 TCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTA 1059

Qy         483 AATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTT 542
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db        1060 AATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTT 1119

Qy         543 ATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGA 602
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db        1120 ATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGA 1179

Qy         603 ATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCC 662
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db        1180 ATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCC 1239

Qy         663 TCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCA 722
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db        1240 TCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCA 1299

Qy         723 AGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCA 782
               IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

```
Db    1300 AGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCA 1359

Qy     783 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 842
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1360 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 1419

Qy     843 CTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAA 902
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1420 CTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAA 1479

Qy     903 GCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAA 962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1480 GCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAA 1539

Qy     963 TCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTT 1022
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1540 TCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTT 1599

Qy    1023 ACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTC 1082
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1600 ACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTC 1659

Qy    1083 CCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGA 1142
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1660 CCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGA 1719

Qy    1143 ATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAAT 1202
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1720 ATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAAT 1779

Qy    1203 TCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCT 1262
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1780 TCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCT 1839

Qy    1263 GTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAA 1322
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1840 GTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAA 1899

Qy    1323 GCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATT 1382
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1900 GCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATT 1959

Qy    1383 GCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACC 1442
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1960 GCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACC 2019

Qy    1443 TCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCT 1502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2020 TCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCT 2079

Qy    1503 GTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCC 1562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2080 GTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCC 2139

Qy    1563 CTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGA 1622
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    2140 CTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTGA 2199
```

```
Qy   1623 CTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCT 1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2200 CTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCT 2259

Qy   1683 CAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGATC 1742
          |||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||||
Db   2260 CAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAATCAAAATTCCAGATC 2319

Qy   1743 GGTTTCCTGAGAGAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAG 1802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2320 GGTTTCCTGAGAGAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAG 2379

Qy   1803 GGAGGGCTGGGGGCGGGGAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAA 1862
          |||||||||||||| ||||||| |||||||| ||||||||||||||||||| |||||
Db   2380 GGAGGGCTGGGGGC-GGGAAAAGAAAAAAAAAAGTCATTTTGAAATTAACCTCCTAAAA 2438

Qy   1863 GGAATTCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAAT 1922
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2439 GGAATTCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAAT 2498

Qy   1923 AGTCCATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGG 1982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2499 AGTCCATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGG 2558

Qy   1983 GTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTA 2042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2559 GTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTA 2618

Qy   2043 CTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGCTG 2102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2619 CTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGCTG 2678

Qy   2103 AGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCT 2162
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2679 AGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCT 2738

Qy   2163 CCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTT 2222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2739 CCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTT 2798

Qy   2223 AATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTT 2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2799 AATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTT 2858

Qy   2283 CACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACA 2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2859 CACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACA 2918

Qy   2343 CAAGCACAACTAAGTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCC 2402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2919 CAAGCACAACTAAGTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCC 2978

Qy   2403 AAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTC 2462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2979 AAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTC 3038
```

```
Qy   2463 TGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTA 2522
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3039 TGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTA 3098

Qy   2523 ACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCA 2582
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3099 ACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCA 3158

Qy   2583 GGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCA 2642
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3159 GGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCA 3218

Qy   2643 AACATTCCTTGTGTTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGAC 2702
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3219 AACATTCCTTGTGTTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGAC 3278

Qy   2703 TCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGA 2762
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   3279 TCCACACTCAGCCTTCTCTACCCTGAATTGAATTATCACCCTTTTCTCCATGTTTTCAGA 3338

Qy   2763 GTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGT 2822
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3339 GTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGT 3398

Qy   2823 GTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTT 2882
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3399 GTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTT 3458

Qy   2883 TTGTTTGTTTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTT 2942
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3459 TTGTTTGTTTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTT 3518

Qy   2943 TAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCG 3002
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3519 TAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCG 3578

Qy   3003 GGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAAC 3062
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3579 GGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAAC 3638

Qy   3063 CCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTT 3122
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3639 CCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTT 3698

Qy   3123 TTCAATTACATCCTGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGT 3182
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3699 TTCAATTACATCCTGACTTGTATAGACACAGCCAGAAAGAAACTGTTAATAGCCATCCGT 3758

Qy   3183 CCATGTAACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAG 3242
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   3759 CCATGTAACTCTGTATTTTACTAAGATACCAATAGCTCTTTCATAGACTTGTGCTACAAG 3818

Qy   3243 AAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAA 3302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3819 AAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAA 3878

Qy   3303 ATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTG 3362
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3879 ATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTG 3938

Qy   3363 TGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAA 3422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||     |  |  ||
Db   3939 TGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATATCTGTAGGAGGAA 3998

Qy   3423 AAAAAAAAAAA 3434
          ||  ||   |   ||
Db   3999 AACAATGACTAA 4010


RESULT 2
AC019335/c
LOCUS       AC019335            182638 bp    DNA      linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-427K3, WORKING DRAFT SEQUENCE,
            18 unordered pieces.
ACCESSION   AC019335
VERSION     AC019335.2  GI:7231064
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182638)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182638)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:6652510.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0427K03
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 87%
            Sequencing vector: plasmid; 13%
            Chemistry: Dye-primer ET; 87% of reads
            Chemistry: Dye-terminator Big Dye; 13% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 174376 bases at least Q40
            Consensus quality: 176799 bases at least Q30
            Consensus quality: 178323 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 180938; sum-of-contigs
            Quality coverage: 4.40 in Q20 bases; agarose-fp
            Quality coverage: 4.46 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
         * consists of 18 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *           1      1193: contig of 1193 bp in length
         *        1194      1293: gap of unknown length
         *        1294      2928: contig of 1635 bp in length
         *        2929      3028: gap of unknown length
         *        3029      5126: contig of 2098 bp in length
         *        5127      5226: gap of unknown length
         *        5227      6713: contig of 1487 bp in length
         *        6714      6813: gap of unknown length
         *        6814      9443: contig of 2630 bp in length
         *        9444      9543: gap of unknown length
         *        9544     13186: contig of 3643 bp in length
         *       13187     13286: gap of unknown length
         *       13287     18101: contig of 4815 bp in length
         *       18102     18201: gap of unknown length
         *       18202     22856: contig of 4655 bp in length
         *       22857     22956: gap of unknown length
         *       22957     31608: contig of 8652 bp in length
         *       31609     31708: gap of unknown length
         *       31709     41973: contig of 10265 bp in length
         *       41974     42073: gap of unknown length
         *       42074     50868: contig of 8795 bp in length
         *       50869     50968: gap of unknown length
         *       50969     59100: contig of 8132 bp in length
         *       59101     59200: gap of unknown length
         *       59201     72609: contig of 13409 bp in length
         *       72610     72709: gap of unknown length
         *       72710     86964: contig of 14255 bp in length
         *       86965     87064: gap of unknown length
         *       87065    104522: contig of 17458 bp in length
         *      104523    104622: gap of unknown length
         *      104623    126360: contig of 21738 bp in length
         *      126361    126460: gap of unknown length
         *      126461    149770: contig of 23310 bp in length
         *      149771    149870: gap of unknown length
         *      149871    182638: contig of 32768 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182638
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-427K3"
     misc_feature    1. .1193
                     /note="assembly_name:Contig4"
     misc_feature    1294. .2928
                     /note="assembly_name:Contig5"
     misc_feature    3029. .5126
                     /note="assembly_name:Contig6"
     misc_feature    5227. .6713
                     /note="assembly_name:Contig7"
```

```
     misc_feature      6814. .9443
                       /note="assembly_name:Contig8"
     misc_feature      9544. .13186
                       /note="assembly_name:Contig9"
     misc_feature      13287. .18101
                       /note="assembly_name:Contig10"
     misc_feature      18202. .22856
                       /note="assembly_name:Contig11
                       clone_end:SP6
                       vector_side:left"
     misc_feature      22957. .31608
                       /note="assembly_name:Contig12"
     misc_feature      31709. .41973
                       /note="assembly_name:Contig13"
     misc_feature      42074. .50868
                       /note="assembly_name:Contig14"
     misc_feature      50969. .59100
                       /note="assembly_name:Contig15"
     misc_feature      59201. .72609
                       /note="assembly_name:Contig16"
     misc_feature      72710. .86964
                       /note="assembly_name:Contig17
                       clone_end:T7
                       vector_side:right"
     misc_feature      87065. .104522
                       /note="assembly_name:Contig18"
     misc_feature      104623. .126360
                       /note="assembly_name:Contig19"
     misc_feature      126461. .149770
                       /note="assembly_name:Contig20"
     misc_feature      149871. .182638
                       /note="assembly_name:Contig21"
BASE COUNT      59178 a   34459 c   33797 g   53497 t   1707 others
ORIGIN
```

```
  Query Match                 59.0%;   Score 2045.4;   DB 2;   Length 182638;
  Best Local Similarity    98.7%;   Pred. No. 0;
  Matches 2072;  Conservative      0;  Mismatches     26;  Indels     1;  Gaps     1;

Qy       1336 TGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 1395
              I IIII  I     IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     179790 TTTCTTTTCAGGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA
179731

Qy       1396 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 1455
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     179730 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG
179671

Qy       1456 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 1515
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     179670 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC
179611

Qy       1516 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 1575
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

```
Db    179610 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT
      179551

Qy      1576 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAG 1635
             ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    179550 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTGACTTCAGGCTCCAG
      179491

Qy      1636 GGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAGCTCTCAATTGAGGAGGA 1695
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179490 GGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAGCTCTCAATTGAGGAGGA
      179431

Qy      1696 GCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAATTCCAGATCGGTTTCCTGAGAG 1755
             |||||||||||||||||||||||||||||||||||||    |||||||||||||||||||||
Db    179430 GCTGGAGAAATGTATCCAGGATTTCCTAAAAATCAAAATTCCAGATCGGTTTCCTGAGAG
      179371

Qy      1756 AAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG 1815
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179370 AAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG
      179311

Qy      1816 CGGGGAAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTT 1875
             |  |||||||  |||||||||  |||||||||||||||||||||| ||||||||||||||||
Db    179310 C-GGGAAAAGAAAAAAAAAAAAGTCATTTTGAAATTAACCTCCTAAAAGGAATTCATATTT
      179252

Qy      1876 TAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATAT 1935
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179251 TAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATAT
      179192

Qy      1936 ACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTC 1995
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179191 ACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTC
      179132

Qy      1996 TAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTACTTCGTCCCATGT 2055
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179131 TAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTACTTCGTCCCATGT
      179072

Qy      2056 GCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGA 2115
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179071 GCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGA
      179012

Qy      2116 GTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGG 2175
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179011 GTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGG
      178952

Qy      2176 GTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTT 2235
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178951 GTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTT
      178892
```

```
Qy      2236  CCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTAT 2295
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178891  CCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTAT
178832

Qy      2296  ATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCACAACTAA 2355
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178831  ATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCACAACTAA
178772

Qy      2356  GTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTC 2415
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178771  GTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTC
178712

Qy      2416  CTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTA 2475
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178711  CTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTA
178652

Qy      2476  AAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTAACATCAAATGACT 2535
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178651  AAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTAACATCAAATGACT
178592

Qy      2536  CTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTAC 2595
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178591  CTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTAC
178532

Qy      2596  TGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTG 2655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178531  TGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTG
178472

Qy      2656  TTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC 2715
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178471  TTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC
178412

Qy      2716  TTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC 2775
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178411  TTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC
178352

Qy      2776  ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT 2835
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178351  ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT
178292

Qy      2836  TGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTTTTGTTTGTTTTAG 2895
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178291  TGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTTTTGTTTGTTTTAG
178232

Qy      2896  AGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTACAAAGC 2955
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178231 AGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTACAAAGC
      178172


Qy      2956 TACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCGGGCTTTGGCTGTG 3015
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178171 TACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCGGGCTTTGGCTGTG
      178112


Qy      3016 CCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAAGGTAAC 3075
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178111 CCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAAGGTAAC
      178052


Qy  .   3076 TGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATTACATCC 3135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178051 TGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATTACATCC
      177992


Qy      3136 TGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTG 3195
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177991 TGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTG
      177932


Qy      3196 TATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGAC 3255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177931 TATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGAC
      177872


Qy      3256 CAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAAT 3315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177871 CAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAAT
      177812


Qy      3316 TAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTAT 3375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177811 TAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTAT
      177752


Qy      3376 GTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAAA 3434
             ||||||||||||||||||||||||||||||||||||||||    |  |   ||||  ||   |   ||
Db    177751 GTGTATCACAGGTAATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA 177693



RESULT 3
AC008716/c
LOCUS       AC008716              184589 bp    DNA       linear    PRI 17-OCT-2001
DEFINITION  Homo sapiens chromosome 5 clone CTB-85P21, complete sequence.
ACCESSION   AC008716
VERSION     AC008716.7  GI:16195190
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184589)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
```

```
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 184589)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 184589)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 184589)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 17, 2001 this sequence version replaced gi:8122137.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.9.
            STS Content:
            WI-13819 G21997.
FEATURES             Location/Qualifiers
     source          1. .184589
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-85P21"
BASE COUNT     60462 a  35809 c  33674 g  54644 t
ORIGIN

    Query Match              58.9%;   Score 2043.4;   DB 9;   Length 184589;
    Best Local Similarity    98.7%;   Pred. No. 0;
    Matches 2071;  Conservative    0;  Mismatches   26;  Indels    2;  Gaps    1;

Qy       1336 TGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 1395
              I IIII I     IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     179121 TTTCTTTTCAGGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA
179062

Qy       1396 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 1455
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     179061 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG
179002

Qy       1456 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 1515
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     179001 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC
178942

Qy       1516 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 1575
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

```
Db     178941 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT
178882

Qy       1576 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAG 1635
              ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db     178881 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTGACTTCAGGCTCCAG
178822

Qy       1636 GGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGA 1695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178821 GGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGA
178762

Qy       1696 GCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAG 1755
              ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||||
Db     178761 GCTGGAGAAATGTATCCAGGATTTCCTAAAAATCAAAATTCCAGATCGGTTTCCTGAGAG
178702

Qy       1756 AAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG 1815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178701 AAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG
178642

Qy       1816 CGGGGAAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTT 1875
              ||||    ||||| |||||||| |||||||||||||||||||||||| ||||||||||||||
Db     178641 CGGG--AAAAGAAAAAAAAAAGTCATTTTGAAATTAACCTCCTAAAAGGAATTCATATTT
178584

Qy       1876 TAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATAT 1935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178583 TAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATAT
178524

Qy       1936 ACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTC 1995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178523 ACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTC
178464

Qy       1996 TAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTACTTCGTCCCATGT 2055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178463 TAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTACTTCGTCCCATGT
178404

Qy       2056 GCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGA 2115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178403 GCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGA
178344

Qy       2116 GTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGG 2175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178343 GTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGG
178284

Qy       2176 GTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTT 2235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178283 GTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTT
178224
```

```
Qy        2236  CCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTAT 2295
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178223  CCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTAT
178164

Qy        2296  ATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCACAACTAA 2355
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178163  ATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCACAACTAA
178104

Qy        2356  GTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTC 2415
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178103  GTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTC
178044

Qy        2416  CTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTA 2475
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178043  CTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTA
177984

Qy        2476  AAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTAACATCAAATGACT 2535
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177983  AAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTAACATCAAATGACT
177924

Qy        2536  CTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTAC 2595
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177923  CTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTAC
177864

Qy        2596  TGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTG 2655
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177863  TGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTG
177804

Qy        2656  TTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC 2715
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177803  TTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC
177744

Qy        2716  TTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC 2775
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177743  TTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC
177684

Qy        2776  ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT 2835
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177683  ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT
177624

Qy        2836  TGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTTTTGTTTGTTTTAG 2895
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     177623  TGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTTTTGTTTGTTTTAG
177564

Qy        2896  AGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTACAAAGC 2955
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177563 AGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTACAAAGC
177504

Qy     2956 TACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCGGGCTTTGGCTGTG 3015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177503 TACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCGGGCTTTGGCTGTG
177444

Qy     3016 CCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAAGGTAAC 3075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177443 CCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAAGGTAAC
177384

Qy     3076 TGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATTACATCC 3135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177383 TGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATTACATCC
177324

Qy     3136 TGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTG 3195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177323 TGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTG
177264

Qy     3196 TATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGAC 3255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177263 TATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGAC
177204

Qy     3256 CAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAAT 3315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177203 CAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAAT
177144

Qy     3316 TAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTAT 3375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177143 TAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTAT
177084

Qy     3376 GTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAA 3434
          |||||||||||||||||||||||||||||||||||||    |  |  ||||  ||  |  ||
Db   177083 GTGTATCACAGGTAATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA 177025


RESULT 4
AC008473/c
LOCUS       AC008473           98360 bp    DNA      linear    PRI 03-OCT-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-375J15, complete sequence.
ACCESSION   AC008473
VERSION     AC008473.6  GI:15887240
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98360)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
```

```
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2   (bases 1 to 98360)
  AUTHORS  DOE Joint Genome Institute.
  TITLE    Direct Submission
  JOURNAL  Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  3   (bases 1 to 98360)
  AUTHORS  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE    Direct Submission
  JOURNAL  Submitted (01-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
REFERENCE  4   (bases 1 to 98360)
  AUTHORS  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE    Direct Submission
  JOURNAL  Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
COMMENT    On Oct 3, 2001 this sequence version replaced gi:9958005.
           Draft Sequence Produced by DOE Joint Genome Institute
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center
           www-shgc.stanford.edu
           Quality: Phrap Quality >=40 99.1% of Sequence;
           Estimated Total Number of Errors is 0.8.
           STS Content:
           SHGC-103102 G57424.
FEATURES             Location/Qualifiers
     source          1. .98360
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-375J15"
BASE COUNT     31520 a  19442 c  18516 g  28882 t
ORIGIN

     Query Match              58.8%;   Score 2038.6;   DB 9;   Length 98360;
     Best Local Similarity    98.5%;   Pred. No. 0;
     Matches 2068;  Conservative    0;  Mismatches   29;  Indels    2;  Gaps    1;

Qy      1336 TGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 1395
             | |||| |     |||||||||||||||||||||||||||||||||||||||||||||||
Db      6692 TTTCTTTTCAGGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 6633

Qy      1396 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 1455
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6632 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 6573

Qy      1456 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 1515
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6572 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 6513

Qy      1516 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 1575
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6512 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 6453

Qy      1576 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAG 1635
```

```
            |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    6452  CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTGACTTCAGGCTCCAG 6393

Qy    1636  GGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGA 1695
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6392  GGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGA 6333

Qy    1696  GCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAG 1755
            ||||||||||||||||||||||||||||||||||||     ||||||||||||||||||||
Db    6332  GCTGGAGAAATGTATCCAGGATTTCCTAAAAATCAAAATTCCAGATCGGTTTCCTGAGAG 6273

Qy    1756  AAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG 1815
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6272  AAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG 6213

Qy    1816  CGGGGAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTT 1875
            ||||    ||||  ||||||||| ||||||||||||||||||||| ||||||||||||||
Db    6212  CGGG--AAAAGAAAAAAAAAAGTCATTTTGAAATTAACCTCCTAAAAGGAATTCATATTT 6155

Qy    1876  TAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATAT 1935
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6154  TAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATAT 6095

Qy    1936  ACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTC 1995
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6094  ACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTC 6035

Qy    1996  TAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTACTTCGTCCCATGT 2055
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6034  TAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTTTACTTCGTCCCATGT 5975

Qy    2056  GCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAGTAGCTGAGAGGCCTTGGGA 2115
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5974  GCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAGTAGCTGAGAGGCCTTGGGA 5915

Qy    2116  GTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGG 2175
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5914  GTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGG 5855

Qy    2176  GTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTT 2235
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5854  GTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTT 5795

Qy    2236  CCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTAT 2295
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5794  CCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTAT 5735

Qy    2296  ATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCACAACTAA 2355
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5734  ATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCACAACTAA 5675

Qy    2356  GTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTC 2415
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5674  GTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTC 5615

Qy    2416  CTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTA 2475
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    5614 CTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTA 5555

Qy    2476 AAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTAACATCAAATGACT 2535
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5554 AAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCTAACATCAAATGACT 5495

Qy    2536 CTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTAC 2595
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5494 CTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTAC 5435

Qy    2596 TGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTG 2655
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5434 TGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTG 5375

Qy    2656 TTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC 2715
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5374 TTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC 5315

Qy    2716 TTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC 2775
            ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    5314 TTCTCTACCCTGAATTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC 5255

Qy    2776 ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT 2835
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5254 ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT 5195

Qy    2836 TGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTTTTGTTTGTTTTAG 2895
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5194 TGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCCTTTTGTTTGTTTTAG 5135

Qy    2896 AGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTACAAAGC 2955
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5134 AGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTACAAAGC 5075

Qy    2956 TACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCGGGCTTTGGCTGTG 3015
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5074 TACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTCGGGCTTTGGCTGTG 5015

Qy    3016 CCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAAGGTAAC 3075
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5014 CCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAAGGTAAC 4955

Qy    3076 TGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATTACATCC 3135
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4954 TGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATTACATCC 4895

Qy    3136 TGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTG 3195
            ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    4894 TGACTTGTATAGACACAGCCAGAAAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTG 4835

Qy    3196 TATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGAC 3255
            |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    4834 TATTTTACTAAGATACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGAC 4775

Qy    3256 CAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAAT 3315
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4774 CAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAAT 4715
```

```
Qy    3316 TAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTAT 3375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4714 TAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTAT 4655

Qy    3376 GTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAA 3434
           |||||||||||||||||||||||||||||||||||||      |  |  ||||  ||  |  ||
Db    4654 GTGTATCACAGGTAATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA 4596
```

RESULT 5
AX405760

```
LOCUS       AX405760                2412 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 175 from Patent WO0222660.
ACCESSION   AX405760
VERSION     AX405760.1  GI:21438959
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
            Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
  TITLE     Novel nucleic acids and polypeptides
  JOURNAL   Patent: WO 0222660-A 175 21-MAR-2002;
            HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .2412
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             1092. .2378
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD34867.1"
                     /db_xref="GI:21438960"
                     /translation="MALSGNCSRYYPREQGSAVPNSFPEVVELNVGGQVYFTRHSTLI
                     SIPHSLLWKMFSPKRDTANDLAKDSKGRFFIDRDGFLFRYILDYLRDRQVVLPDHFPE
                     KGRLKREAEYFQLPDLVKLLTPDEIKQSPDEFCHSDFEDASQGSDTRICPPSSLLPAD
                     RKWGFITVGYRGSCTLGREGQADAKFRRVPRILVCGRISLAKEVFGETLNESRDPDRA
                     PERYTSRFYLKFKHLERAFDMLSECGFHMVACNSSVTASFINQYTDDKIWSSYTEYVF
                     YREPSRWSPSHCDCCCKNGKGDKEGESGTSCNDLSTSSCDSQSEASSPQETVICGPVT
                     RQTNIQTLDRPIKKGPVQLIQQSEMRRKSDLLRTLTSGSRESNMSSKKKAVKEKLSIE
                     EELEKCIQDFLKIKIPDRFPERKHPWQSELLRKYHL"
BASE COUNT      638 a     585 c     551 g     638 t
ORIGIN
```

```
     Query Match            47.3%;   Score 1640.8;   DB 6;   Length 2412;
     Best Local Similarity  99.6%;   Pred. No. 0;
     Matches 1645;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;
```

```
Qy     183 ACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGA 242
           || || |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 ACTAAGACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGA 819

Qy     243 AGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCA 302
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      820 AGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCA 879

Qy      303 TCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGAT 362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      880 TCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGAT 939

Qy      363 TTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTT 422
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      940 TTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTT 999

Qy      423 TCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTA 482
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1000 TCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTA 1059

Qy      483 AATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTT 542
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1060 AATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTT 1119

Qy      543 ATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGA 602
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1120 ATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGA 1179

Qy      603 ATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCC 662
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1180 ATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCC 1239

Qy      663 TCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCA 722
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1240 TCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCA 1299

Qy      723 AGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCA 782
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1300 AGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCA 1359

Qy      783 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 842
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1360 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 1419

Qy      843 CTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAA 902
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1420 CTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAA 1479

Qy      903 GCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAA 962
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1480 GCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAA 1539

Qy      963 TCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTT 1022
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1540 TCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTT 1599

Qy     1023 ACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTC 1082
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1600 ACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTC 1659

Qy     1083 CCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGA 1142
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     1660 CCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGA 1719

Qy     1143 ATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAAT 1202
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1720 ATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAAT 1779

Qy     1203 TCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCT 1262
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1780 TCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCT 1839

Qy     1263 GTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAA 1322
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1840 GTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAA 1899

Qy     1323 GCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATT 1382
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1900 GCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATT 1959

Qy     1383 GCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACC 1442
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1960 GCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACC 2019

Qy     1443 TCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCT 1502
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2020 TCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCT 2079

Qy     1503 GTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCC 1562
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2080 GTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCC 2139

Qy     1563 CTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGA 1622
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     2140 CTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTGA 2199

Qy     1623 CTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAGCTCT 1682
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2200 CTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAGCTCT 2259

Qy     1683 CAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGATC 1742
            ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||||||
Db     2260 CAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAATCAAAATTCCAGATC 2319

Qy     1743 GGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAG 1802
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2320 GGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAG 2379

Qy     1803 GGAGGGCTGGGGGCGGGGAAAAAAAAAAAAAA 1834
            ||||||||||||||||||| ||| |||||||||
Db     2380 GGAGGGCTGGGGGCGGGAAAAGAAAAAAAAA 2411


RESULT 6
AC008652/c
LOCUS       AC008652              171949 bp    DNA      linear    PRI 31-JUL-2001
DEFINITION  Homo sapiens chromosome 5 clone CTB-18F1, complete sequence.
ACCESSION   AC008652
VERSION     AC008652.6  GI:15042788
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171949)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171949)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 171949)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 31, 2001 this sequence version replaced gi:9256021.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
FEATURES             Location/Qualifiers
     source          1. .171949
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-18F1"
BASE COUNT     54666 a  33484 c  31658 g  52141 t
ORIGIN

    Query Match              33.5%;   Score 1161.8;   DB 9;   Length 171949;
    Best Local Similarity    99.8%;   Pred. No. 1.4e-240;
    Matches 1163;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;


Qy        187 ATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGAAGAT 246
              I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     171266 AGACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGAAGAT
171207


Qy        247 TGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCATCCA 306
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     171206 TGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCATCCA
171147


Qy        307 GGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGATTTCT 366
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db     171146 GGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGATTTCT
171087


Qy        367 TGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTTTCCC 426
              IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

```
Db      171086 TGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTTTCCC
171027

Qy         427 TTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTAAATC 486
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      171026 TTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTAAATC
170967

Qy         487 AAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTTATTA 546
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170966 AAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTTATTA
170907

Qy         547 TCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGT 606
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170906 TCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGT
170847

Qy         607 CGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCT 666
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170846 CGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCT
170787

Qy         667 GTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGG 726
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170786 GTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGG
170727

Qy         727 AAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGA 786
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170726 AAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGA
170667

Qy         787 CAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGA 846
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170666 CAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGA
170607

Qy         847 ATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGCCC 906
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170606 ATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGCCC
170547

Qy         907 AGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAATCTG 966
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170546 AGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAATCTG
170487

Qy         967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170486 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG
170427

Qy        1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170426 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG
170367
```

```
Qy        1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170366 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA
170307


Qy        1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170306 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA
170247


Qy        1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170246 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA
170187


Qy        1267 CTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAAGCTA 1326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170186 CTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAAGCTA
170127


Qy        1327 CACTGAATATGTCTTCTACCGTGAG 1351
               ||||||||||||||||||||||| ||
Db      170126 CACTGAATATGTCTTCTACCGTAAG 170102



RESULT 7
AC008383
LOCUS       AC008383              209114 bp    DNA        linear   PRI 01-MAY-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-222O22, complete sequence.
ACCESSION   AC008383
VERSION     AC008383.8  GI:13899395
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209114)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209114)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 209114)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:13699337.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
```

FEATURES   Location/Qualifiers
     source   1. .209114
                         /organism="Homo sapiens"
                         /mol_type="genomic DNA"
                         /db_xref="taxon:9606"
                         /chromosome="5"
                         /clone="CTC-222O22"
BASE COUNT   65378 a   39329 c   40172 g   64235 t
ORIGIN

  Query Match   33.5%;   Score 1161.8;   DB 9;   Length 209114;
  Best Local Similarity   99.8%;   Pred. No. 1.4e-240;
  Matches 1163; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   187 ATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGAAGAT 246
                I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168231 AGACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGAAGAT
168290

Qy   247 TGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCATCCA 306
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168291 TGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCATCCA
168350

Qy   307 GGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGATTTCT 366
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168351 GGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGATTTCT
168410

Qy   367 TGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTTTCCC 426
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168411 TGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTTTCCC
168470

Qy   427 TTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTAAATC 486
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168471 TTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTAAATC
168530

Qy   487 AAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTTATTA 546
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168531 AAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTTATTA
168590

Qy   547 TCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGT 606
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168591 TCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGT
168650

Qy   607 CGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCT 666
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db   168651 CGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCT
168710

Qy   667 GTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGG 726
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII

```
Db      168711 GTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGG
        168770

Qy         727 AAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGA 786
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      168771 AAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGA
        168830

Qy         787 CAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGA 846
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      168831 CAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGA
        168890

Qy         847 ATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGCCC 906
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      168891 ATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGCCC
        168950

Qy         907 AGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAATCTG 966
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      168951 AGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAATCTG
        169010

Qy         967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      169011 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG
        169070

Qy        1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      169071 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG
        169130

Qy        1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      169131 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA
        169190

Qy        1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      169191 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA
        169250

Qy        1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      169251 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA
        169310

Qy        1267 CTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAAGCTA 1326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      169311 CTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAAGCTA
        169370

Qy        1327 CACTGAATATGTCTTCTACCGTGAG 1351
               |||||||||||||||||||||| ||
Db      169371 CACTGAATATGTCTTCTACCGTAAG 169395
```

```
RESULT 8
AC127249/c
LOCUS       AC127249              135132 bp    DNA     linear   HTG 12-JUN-2003
DEFINITION  Mus musculus chromosome UNK clone RP24-475B8, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC127249
VERSION     AC127249.3  GI:31621481
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 135132)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135132)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 135132)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jun 12, 2003 this sequence version replaced gi:21886968.

            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BB0475B08
            -----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 133026 bases at least Q40
            Consensus quality: 133293 bases at least Q30
            Consensus quality: 133464 bases at least Q20
            ---------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    8786: contig of 8786 bp in length
            *     8787    8886: gap of unknown length
```

```
                 *      8887      23581: contig of 14695 bp in length
                 *     23582      23681: gap of unknown length
                 *     23682      43616: contig of 19935 bp in length
                 *     43617      43716: gap of unknown length
                 *     43717     135132: contig of 91416 bp in length.
FEATURES             Location/Qualifiers
     source          1. .135132
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-475B8"
     misc_feature    1. .8786
                     /note="assembly_name:Contig21"
     misc_feature    8887. .23581
                     /note="assembly_name:Contig22"
     misc_feature    23682. .43616
                     /note="assembly_name:Contig23"
     misc_feature    43717. .135132
                     /note="assembly_name:Contig24"
BASE COUNT    44115 a  26731 c  25600 g  38364 t    322 others
ORIGIN
```

```
  Query Match              24.6%;   Score 853.2;  DB 2;   Length 135132;
  Best Local Similarity    70.7%;   Pred. No. 7.5e-174;
  Matches 1473;  Conservative    0;  Mismatches  523;  Indels   86;  Gaps    22;


Qy       1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
              ||||||||||| | ||||| |||| || || || ||||||||||||||||||||| || |
Db     124743 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG
124684


Qy       1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
              |||    |||| ||||||||||||| || ||||||||||| ||||| || ||||| |||||||
Db     124683 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC
124627


Qy       1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
              |||| ||||||||||||||| |||||||||| |||||||| || || || |||||||| ||
Db     124626 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA
124567


Qy       1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
              ||||||||||||||||| || |||||||||| || || || |||||||||||||||||||||
Db     124566 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG
124507


Qy       1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
              ||||| ||||||||||| ||| | ||||||| |||||| ||||||||||||||||| |||||||
Db     124506 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA
124447


Qy       1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
              | ||||||||||| ||||||| || ||||||||||| || ||||| |||||||||||||
Db     124446 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT
124387
```

```
Qy        1707 GTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
               ||||||||||||| | || | ||||||||||||||| || |||||| ||||||||||||
Db      124386 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAAACATCCTT
124327

Qy        1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
               |||| ||||||||||||| ||||||||||||||||| ||| || | ||
Db      124326 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGGAGGGCTGTGG----------
124278

Qy        1827 AAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
                |    | || |||||||| |||||| ||||||| ||||| |||||||||||||
Db      124277 ----GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAGACATATGTTAAAGGAAAAA
124222

Qy        1887 AATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTA 1946
               | |||||||| | | ||||||| ||||| |||||| ||||||||||| ||||||||||||
Db      124221 TA-ACAACTAACGGTCCACATTTGTTAGATCACAAT-GTCCATTGATGTACTACTGCCTA
124164

Qy        1947 CTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA 2006
               |||| |||||| |||||||||| |||||||||||||||||||||||||||||||||||||
Db      124163 CTTTGCCTAGCTCACCTTAACGTGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA
124104

Qy        2007 GTACAAGAAAATCTTTTTTAGTTATTTG---TTTGTTTACTTCGTCCCATGTGCTAACTA 2063
                | || |   || | || |||| ||| ||||||||| ||||||||||||| | |
Db      124103 CCAGAAACGAGCTCTTAGTTGTCCTTTGTCTTTTATTTACTTGGTCCCATGTGCTGAGAA
124044

Qy        2064 TCTT-ATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCATTT 2122
               |||| | ||| || ||| ||||||||||| ||||||| | |||| |||||| |||||
Db      124043 TCTTAAGATACAACAAGAACCAGCTACGTGTGAGTAGCTCACAGGCTTTGGGAATCATTG
123984

Qy        2123 ATCCCAAACTGGGTTTTTT-----------------------CTCTCATCCTTCTAC 2156
               |||||||| |||||||||                        ||||||| |||| |
Db      123983 ATCCCAAACCAGGTTTTTTTGTTTTGTTTTGTTTTGTTTTGTTTACTCTCATTTTTCTGC
123924

Qy        2157 CTCCCTCCTTTGA--ATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTG 2214
               |||| || ||||| | || | | || ||| |||| | ||| |||| || |||
Db      123923 CTCCTCCCCTTGACCAAGAATGGACAGTTGAAGGAGATATAACCCGGTGGCTTATGTTAA
123864

Qy        2215 GAATTTTTAATTTTGGTTTTTTCCT---TTTGTTTATGGGGTT--GGGGGGAATGGCAGAT 2269
               ||| || | |||| ||| | | ||||||||||||||||| |||| |||||||| ||
Db      123863 GAAATTATCCTTTTCCCTTTCCTTTTGTTTGTTTATGGGGTTGAGGGGAGAATGGCAAAT
123804

Qy        2270 TTATATGACTTTTCACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAG 2329
               || ||||| ||||||||| |||||||| |||||||||| || ||||||||| ||||||||||
Db      123803 TTGTATGATTTTTCACTAAAATCTCTATGTGCCAGGTTCTATTGACTTTGTATGCATGAG
123744

Qy        2330 TATTTGTGCAACACAAGCA-CAACTAAGTATGTATATACACATGACGCACACGATGCCAG 2388
               ||| || |||    || | |    | | |||||| ||||          |||    ||
```

```
Db     123743 CGTTTCTGACACAAGCACAGTATATGTCTGTATATATGCACAAAGAATGCACACGACCTA
       123684

Qy       2389 GGCCTAGACCTCCCAAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCA 2448
              || |  | |  |  | ||||||   || | ||  |||| |||| ||   ||| | ||||
Db     123683 GGGCCTGGACAGCAGAGGGCTAACATCTTACTATCAGCTGCCC-CTACAAGAGCACTTCA
       123625

Qy       2449 GATGGATGAGCTTCTGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTT 2508
              ||     ||| |||| || ||| |||||| || |||| ||||||||||||||||   |||
Db     123624 GACAACCAAGCCTCTGCCTATTTATTAAAACCCTCCTGGGCAGATTTCCCAGCCTCCCTT
       123565

Qy       2509 CACAACACTTTCT-----AACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTAT 2563
              ||        ||     ||  | ||   | |  ||||||||||||||| | || |||||
Db     123564 GGCAGGCAGCACTTTCTAAAGCTGAATAGGCCCCCATCATCAACAAATTCTCTTTCTTAT
       123505

Qy       2564 TGTGAAATTAATACCCTCAGGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGA 2623
              | ||||||| |||||| |||||||| |||||||| ||     |  | || |||||  |||
Db     123504 TTTGAAATAAATACCCACAGGCTCCTTTGATTTATTATGTTCTTTCCCTACATTAGGAGC
       123445

Qy       2624 GGATGAGGAGAGCTGGTCAAACATTCCTTGTGTTAAA---AAAATCAAACATTCATATCC 2680
              | ||||  ||||| ||| ||| |||  | |||||||| || |  || |||||| |||||||
Db     123444 TGGTGAGATGAGCTAGTCTAACCCTGTTTGTGTTAACAGACAAGCAAACAGTCATATCC
       123385

Qy       2681 ACAAAATTTTCTGCTAAATGACTCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCA 2740
              |||||        || | || ||   | |||||||||||||||   || |   ||| ||||
Db     123384 ACAAACAGAG-TGTTGAAAGATCTCGCACTCAGCCTTCTCCGTTCTAATTAGAACAATCA
       123326

Qy       2741 CCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCA 2800
              || || |||  || |  |    || |||  ||  || | ||| || | | || |
Db     123325 CCATTCTCTAGCTGACTCAGAGTTTTAAACTTGCCCACATTTTATTAACAAGGCCTTTGA
       123266

Qy       2801 CATAATCCACATTAAGTTCTGTGTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGT 2860
              |||||||     || |       |||    |||||| | ||| | |||||| || |
Db     123265 TATAATCCAGGCAAATTCTCTGCCTCCCTATGGGTTGTGAAGCTACGAACAACACCCAAT
       123206

Qy       2861 ACTTGAATAAGGGTCCGGCCTTTTGTT-TGTTTTAGAGAAAGTTGTATTCCACACACAAC 2919
              ||||| | |  ||| ||||| ||| || ||   | | |  | | || ||     |
Db     123205 GATTGAAAATGCATCCAGCCTTCCGTTCCCTTGTTTTAGAGGATTTGTGCCCCAACATAT
       123146

Qy       2920 CTAATAATTTCTTATAAAAATTTTAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCG 2979
              |  |||||||| ||| |||  |||| ||||||   |  | | | || |||||||||||||
Db     123145 GTCTAAATTTCTCATACAAACTTTACACTACACCTGTTTACTGTTGCTTGCTTGTAGCCA
       123086

Qy       2980 TTTTTGTTTGCCTTTGGGATTC-GGGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGT 3038
              ||||  |  |||||||| || | ||| |||||||||||||| || | ||| | |||| ||| |
Db     123085 GGTTTGGATAACTTTGGAATCCTGGGGTTTGGCTGTGGCCCTACTACGGTTTAGTTGTAT
       123026
```

```
Qy        3039 CATTTTTATGATGTCTGTAACAACCCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTC 3098
               |||| ||  |||||||||  ||||||  |||  ||| | |  | |||||    ||
Db      123025 AATTTCTACAGTGTCTGTAATGACCCAAGTGGGTGGCTGGAACATAAAAGTTACTAATTT
122966


Qy        3099 AGATTTCTAAATGAAACTATCTTTTTCAATTACATCCTGACTTGTATAGACACAGCCAAA 3158
               ||| ||||         ||||| ||       ||||||| ||| |||| ||    | ||
Db      122965 GATTTTTAAA---------CTTTTAAAAAATATTCCTGACCTGTGTAGATACCATCCAA
122915


Qy        3159 AAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGC 3218
               ||||||||||||| | || ||||||| | ||||| || |||||| | ||| |||
Db      122914 AAGAAACTGTTAACA-CTGTCTATCCATGTGATTCTGTCTTCTACTAATCTTCCAGTAGT
122856


Qy        3219 TCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTT-ATTTTCAGCATTCCT 3277
               |||||  |  || ||||||| ||||| ||  ||||||| | |||| |||||||||||||||
Db      122855 TCTTTTGTTTAC-TGTGCTAAAAGAATGTCCAAAGACAACTTTTAATTTTCAGCATTCCT
122797


Qy        3278 CATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACT 3337
               ||| |||  ||||||||| ||||     ||   | ||      ||||||||||| |||
Db      122796 CATACATCCAAGTGGTAACTGAAAAGATGATTTATCACTA----GTGTGTGCCAAGAACT
122741


Qy        3338 CCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGT 3379
               |||| ||| || |  || ||||||| ||  ||||||| ||| |
Db      122740 CCTATTTTTTTGTTGTGTGTGTGTCTGTGTGTGTGTGTTT 122699




RESULT 9
AC114984
LOCUS       AC114984              186417 bp    DNA     linear   HTG 05-JUN-2003
DEFINITION  Mus musculus clone RP23-248F9, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   AC114984
VERSION     AC114984.6  GI:30984634
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 186417)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-248F9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186417)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
```

```
              Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
              MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
              McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
              Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
              Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
              Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
              Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
              Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
              Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
              Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
              Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
              Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE       Direct Submission
  JOURNAL     Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE     3  (bases 1 to 186417)
  AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
              Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
              Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
              Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
              Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
              Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
              Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
              Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
              Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
              Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
              Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
              Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
              Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
              O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
              Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
              Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
              Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
              Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
              Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
              Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE       Direct Submission
  JOURNAL     Submitted (05-JUN-2003) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT       On May 22, 2003 this sequence version replaced gi:30023906.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              -------------- Genome Center
                  Center: Whitehead Institute/ MIT Center for Genome Research
                  Center code: WIBR
                  Web site: http://www-seq.wi.mit.edu
                  Contact: sequence_submissions@genome.wi.mit.edu
              -------------- Project Information
                  Center project name: L19035
                  Center clone name: 248_F_9
              --------------.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 6 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
```

```
        *  runs of N, but the exact sizes of the gaps are unknown.
        *  This record will be updated with the finished sequence
        *  as soon as it is available and the accession number will
        *  be preserved.
        *       1     32842: contig of 32842 bp in length
        *    32843     32942: gap of 100 bp
        *    32943     61950: contig of 29008 bp in length
        *    61951     62050: gap of 100 bp
        *    62051     70963: contig of 8913 bp in length
        *    70964     71063: gap of 100 bp
        *    71064     74030: contig of 2967 bp in length
        *    74031     74130: gap of 100 bp
        *    74131    162809: contig of 88679 bp in length
        *   162810    162909: gap of 100 bp
        *   162910    186417: contig of 23508 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186417
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-248F9"
                     /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT     54373 a   36475 c   36595 g   58359 t      615 others
ORIGIN

  Query Match              24.6%;   Score 853.2;   DB 2;   Length 186417;
  Best Local Similarity    70.7%;   Pred. No. 7.5e-174;
  Matches 1473;  Conservative    0;  Mismatches  523;   Indels   86;   Gaps   22;


Qy      1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
             |||||||||| | ||||| |||| || || || ||||||||||||||||||||| || |
Db    172660 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG
172719


Qy      1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
             |||    |||| ||||||||||| || ||||||||||| ||||| || ||||| |||||||
Db    172720 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC
172776


Qy      1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| ||||||||||||||| ||||||||||| ||||||||| || || || |||||||| ||
Db    172777 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA
172836


Qy      1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
             ||||||||||||||||| || ||||||||||| || || || |||||||||||||||||||
Db    172837 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG
172896


Qy      1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
             ||||| |||||||||| ||| | ||||||| |||||| |||||||||||||| |||||||
Db    172897 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA
172956


Qy      1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
             | ||||||||||| ||||||| || ||||||||||| || ||||| ||||||||||||||
```

```
Db     172957 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT
173016

Qy       1707 GTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAACATCCTT 1766
              |||||||||||||| |  || |  ||||||||||||||| || ||||||  |||||||||
Db     173017 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAACATCCTT
173076

Qy       1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
              |||| ||||||||||| ||||||||||||||||| ||| ||  |  | ||
Db     173077 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGAGGGCTGTGG----------
173125

Qy       1827 AAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
                 |   | || ||||||||| ||||||    ||||||| |||||  ||||||||||||
Db     173126 ----GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAGACATATGTTAAAGGAAAAA
173181

Qy       1887 AATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTA 1946
              | ||||||||| | | |||||||| ||||| |||||| |||||||||  |||||||||||
Db     173182 TA-ACAACTAACGGTCCACATTTGTTAGATCACAAT-GTCCATTGATGTACTACTGCCTA
173239

Qy       1947 CTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA 2006
              |||| ||||| |||||||||| |||||||||||||||||||||||||||||||||||||||
Db     173240 CTTTGCCTAGCTCACCTTAACGTGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA
173299

Qy       2007 GTACAAGAAAATCTTTTTTAGTTATTTG---TTTGTTTACTTCGTCCCATGTGCTAACTA 2063
              |  ||   |    ||  | ||  | || ||||   ||| ||||||||||||||||  |  |
Db     173300 CCAGAAACGAGCTCTTAGTTGTCCTTTGTCTTTTATTTACTTGGTCCCATGTGCTGAGAA
173359

Qy       2064 TCTT-ATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCATTT 2122
              |||| |  ||| ||  ||| ||||||||||   |||||||| | |||| |||||| |||||
Db     173360 TCTTAAGATACAACAAGAACCAGCTACGTGTGAGTAGCTCACAGGCTTTGGGAATCATTG
173419

Qy       2123 ATCCCAAACTGGGTTTTTT------------------------CTCTCATCCTTCTAC 2156
              |||||||| ||||||||||                        |||||||  |||| |
Db     173420 ATCCCAAACCAGGTTTTTTTGTTTTGTTTTGTTTTGTTTTGTTTACTCTCATTTTTCTGC
173479

Qy       2157 CTCCCTCCTTTGA--ATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTG 2214
              |||| || |||| |  || | ||  ||| | ||||| |  |||| |||| ||| || |||
Db     173480 CTCCTCCCCTTGACCAAGAATGGACAGTTGAAGGAGATATAACCCGGTGGCTTATGTTAA
173539

Qy       2215 GAATTTTTAATTTTGGTTTTTCCT---TTTGTTTATGGGGTT--GGGGGGAATGGCAGAT 2269
              ||| || | |||| |||| |  |||  ||||||||||||||  ||||  ||||||||| ||
Db     173540 GAAATTATCCTTTTCCCTTTCCTTTTGTTTGTTTATGGGGTTGAGGGGGAGAATGGCAAAT
173599

Qy       2270 TTATATGACTTTTCACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAG 2329
              || |||||| |||||||| |||||| ||||||||||  || |||||||  |||||||||||
Db     173600 TTGTATGATTTTTCACTAAAATCTCTATGTGCCAGGTTCTATTGACTTTGTATGCATGAG
173659
```

```
Qy        2330 TATTTGTGCAACACAAGCA-CAACTAAGTATGTATATACACATGACGCACACGATGCCAG 2388
               ||| || |||    || | |   | | ||||| ||||      ||| ||
Db      173660 CGTTTCTGACACAAGCACAGTATATGTCTGTATATATGCACAAAGAATGCACACGACCTA
173719

Qy        2389 GGCCTAGACCTCCCAAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCA 2448
               || | | | | | |||||| || | || |||| |||| || ||| | ||||
Db      173720 GGGCCTGGACAGCAGAGGGCTAACATCTTACTATCAGCTGCCC-CTACAAGAGCACTTCA
173778

Qy        2449 GATGGATGAGCTTCTGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTT 2508
               ||    ||| |||| || ||| |||||| || |||| ||||||||||||||||||   |||
Db      173779 GACAACCAAGCCTCTGCCTATTTATTAAAACCCTCCTGGGCAGATTTCCCAGCCTCCCTT
173838

Qy        2509 CACAACACTTTCT-----AACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTAT 2563
               ||         ||      || | || | | |||||||||||||||||| | || |||||
Db      173839 GGCAGGCAGCACTTTCTAAAGCTGAATAGGCCCCCATCATCAACAAATTCTCTTTCTTAT
173898

Qy        2564 TGTGAAATTAATACCCTCAGGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGA 2623
               | |||||| ||||||| ||||||||| |||| ||     | | || ||||| || ||||| |||
Db      173899 TTTGAAATAAATACCCACAGGCTCCTTTGATTTATTATGTTCTTTCCCTACATTAGGAGC
173958

Qy        2624 GGATGAGGAGAGCTGGTCAAACATTCCTTGTGTTAAA---AAAATCAAACATTCATATCC 2680
               | |||| ||||| ||| ||| ||| |||||||| || | || ||||||| |||||||||
Db      173959 TGGTGAGATGAGCTAGTCTAACCCTGTTTGTGTTTAACAGACAAGCAAACAGTCATATCC
174018

Qy        2681 ACAAAATTTTCTGCTAAATGACTCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCA 2740
               |||||       || | || ||   | |||||||||||||||||||   || |   |||| ||||
Db      174019 ACAAACAGAG-TGTTGAAAGATCTCGCACTCAGCCTTCTCCGTTCTAATTAGAACAATCA
174077

Qy        2741 CCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCA 2800
               || || |||     || |     ||    ||| ||     ||| ||     | | || |
Db      174078 CCATTCTCTAGCTGACTCAGAGTTTTAAACTTGCCCACATTTTATTAACAAGGCCTTTGA
174137

Qy        2801 CATAATCCACATTAAGTTCTGTGTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGT 2860
               |||||||     || |        |||       |||||| | ||| | ||||||| || |
Db      174138 TATAATCCAGGCAAATTCTCTGCCTCCCTATGGGTTGTGAAGCTACGAACAACACCCAAT
174197

Qy        2861 ACTTGAATAAGGGTCCGGCCTTTTGTT-TGTTTTAGAGAAAGTTGTATTCCACACACAAC 2919
               ||||| | |   ||| ||||| |||    || |        | | | | | || ||       |
Db      174198 GATTGAAAATGCATCCAGCCTTCCGTTCCCTTGTTTTAGAGGATTTGTGCCCCAACATAT
174257

Qy        2920 CTAATAATTTCTTATAAAAATTTTAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCG 2979
               |       ||||||| ||| ||| |||| |||||||        | | | || |||||||||||||||
Db      174258 GTCTAAATTTCTCATACAAACTTTACACTACACCTGTTTACTGTTGCTTGCTTGTAGCCA
174317

Qy        2980 TTTTTGTTTGCCTTTGGGATTC-GGGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGT 3038
```

```
              |||| | |||||| || | ||| |||||||||| || | ||| | ||||| ||| |
Db    174318 GGTTTGGATAACTTTGGAATCCTGGGGTTTGGCTGTGGCCCTACTACGGTTTAGTTGTAT
174377


Qy      3039 CATTTTTATGATGTCTGTAACAACCCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTC 3098
              |||| ||    |||||||||| ||||||   ||| ||| || | | ||||||      ||
Db    174378 AATTTCTACAGTGTCTGTAATGACCCAAGTGGGTGGCTGGAACATAAAAGTTACTAATTT
174437


Qy      3099 AGATTTCTAAATGAAACTATCTTTTTCAATTACATCCTGACTTGTATAGACACAGCCAAA 3158
              ||| ||||          ||||| ||       ||||||| ||| |||| ||   | ||
Db    174438 GATTTTTTAAA---------CTTTTAAAAAATATTCCTGACCTGTGTAGATACCATCCAA
174488


Qy      3159 AAGAAACTGTTAATAGCCATCCGTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGC 3218
              |||||||||||||| | | || ||||||| | ||||| || ||||||| | ||| |||
Db    174489 AAGAAACTGTTAACA-CTGTCTATCCATGTGATTCTGTCTTCTACTAATCTTCCAGTAGT
174547


Qy      3219 TCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTT-ATTTTCAGCATTCCT 3277
              |||||  |  || |||||||| ||||| ||  |||||| | |||| |||||||||||||||
Db    174548 TCTTTTGTTTAC-TGTGCTAAAAGAATGTCCAAAGACAACTTTTAATTTTCAGCATTCCT
174606


Qy      3278 CATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACT 3337
              ||| |||    |||||||||| ||||     ||    | ||      |||||||||| |||
Db    174607 CATACATCCAAGTGGTAACTGAAAAGATGATTTATCACTA----GTGTGTGCCAAGAACT
174662


Qy      3338 CCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGT 3379
              |||| ||| || | | || ||||||| || ||||||| ||| |
Db    174663 CCTATTTTTTTGTTGTGTGTGTGTCTGTGTGTGTGTGTGTTT 174704



RESULT 10
AC117867/c
LOCUS       AC117867               242679 bp    DNA        linear    HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-376I9, *** SEQUENCING IN PROGRESS
            ***, 11 unordered pieces.
ACCESSION   AC117867
VERSION     AC117867.4  GI:23618130
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242679)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
```

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 242679)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (11-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 242679)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Oct 9, 2002 this sequence version replaced gi:21746224. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GTZA
    Center clone name: CH230-376I9
-------------- Summary Statistics
    Assembly program: Phrap; version 0.990329
    Consensus quality: 188097 bases at least Q40
    Consensus quality: 190770 bases at least Q30
    Consensus quality: 192614 bases at least Q20
    Estimated insert size: 191086; sum-of-contigs estimation
    Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1   185634: contig of 185634 bp in length
*   185635   185734: gap of unknown length
*   185735   189337: contig of 3603 bp in length
*   189338   189437: gap of unknown length
*   189438   198672: contig of 9235 bp in length
*   198673   198772: gap of unknown length
*   198773   202047: contig of 3275 bp in length
*   202048   202147: gap of unknown length
*   202148   203477: contig of 1330 bp in length
*   203478   203577: gap of unknown length
*   203578   204617: contig of 1040 bp in length
*   204618   204717: gap of unknown length
*   204718   205848: contig of 1131 bp in length
*   205849   205948: gap of unknown length
*   205949   207199: contig of 1251 bp in length
*   207200   207299: gap of unknown length
*   207300   208535: contig of 1236 bp in length
*   208536   208635: gap of unknown length
*   208636   211249: contig of 2614 bp in length
*   211250   211349: gap of unknown length
*   211350   242679: contig of 31330 bp in length.

```
FEATURES             Location/Qualifiers
     source          1. .242679
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-376I9"
     misc_feature    1. .1668
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    1914. .2738
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXAWM53TV"
     misc_feature    8953. .9899
                     /note="clone_boundary
                     clone_end:T7
                     site:MboI
                     end_sequence:RXAWM53TJ"
     misc_feature    185735. .187055
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    188311. .189337
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    198773. .199845
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT     59286 a  39026 c  39489 g  55897 t  48981 others
ORIGIN

   Query Match              23.6%;   Score 819.4;   DB 2;   Length 242679;
   Best Local Similarity    83.8%;   Pred. No. 1.5e-166;
   Matches  995;  Conservative    0;  Mismatches  166;  Indels   26;  Gaps     5;


Qy        187 ATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGAAGAT 246
              | || ||| |||| ||||||||| | ||||||| | | |     || | | ||||||||
Db     239544 AGACAGACCTCTGGGTAACTGGGCATTTGGCCTTCTTGCCTACCGAACCTAATGGAAGAC
239485


Qy        247 TGGATATAGACGAGTTGATTAT---ATTTTATGAAGTAGCAGCTCACTACCATCCACCAT 303
              || || ||||| || ||||||    ||| | ||||||| |||||||||||| ||| | |
Db     239484 TGAATGTAGACAAGCTGATTATTACATTCTGTGAAGTAAGAGCTCACTACCAGCCAGCTT
239425


Qy        304 CCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGATT 363
                |   | |    | | | | | |||||| |||||||||||||||||| || | |||| ||
Db     239424 TAAACTTCTCGTCCA-GTGTCCAGCATTTCTTCACCTGTGGACTCTCCTAAAATTTGCTT
239366


Qy        364 TCTTGGG----------------GGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAA 406
              | |||||             || || || || ||||||||||| |||    ||
Db     239365 TATTGGGAAAATAAAAAACAAAATAAAACAAAACCTGGGTAAGAGGAGGTAATTAACAAA
239306


Qy        407 --TAAGTTAGCATCCTTTTCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTC 464
                || |       ||||||||| | ||||| ||| ||| |||||| | |||||| || |
```

```
Db      239305  AGAAAAGTTAGTACCTTTTCCTTACCTTACCAGTGGATGCAAAGGCCAGGGCTGTAACCC
239246

Qy         465  CATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGA 524
                ||||||||||||||| |  || |      |  ||| |||||||||||||||||||||||||
Db      239245  AGTTGGATTGCACCTTAAGTTCCA---GGAAGCTGCAGAAGAAAGGGACAATGGCTCTGA
239189

Qy         525  GTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCC 584
                ||||  |||||||| ||||||||||||||| || ||||||| | || |||||||||| ||||
Db      239188  GTGGGAACTGTAGCCGTTATTATCCTCGGGACCAAGGGGCTGCTGTTCCCAACTCTTTCC
239129

Qy         585  CTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGA 644
                |||| || |||||||||||||| ||||| || ||||| ||||||||||||||||||||| |
Db      239128  CTGAAGTCATAGAGCTGAATGTTGGGGGCCAGGTTTACTTTACTCGCCATTCCACATTAA
239069

Qy         645  TAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATG 704
                |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||| ||||| |
Db      239068  TAAGTATCCCCCATTCTCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACTGCTAACG
239009

Qy         705  ATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTT 764
                ||||||||||||||||||||||||||||||||||| |||||||||||| | ||  ||||||||| .
Db      239008  ATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATCGACAGAGATGGCTTTCTGTTCCGTT
238949

Qy         765  ATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAG 824
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      238948  ATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAGAG
238889

Qy         825  GAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCC 884
                |||| |||||||| |||||||| || ||||||||||| ||| | |||||||||||| |||
Db      238888  GAAGGCTGAAAAGAGAAGCTGAGTATTTCCAGCTCCCTGACCTCGTCAAACTCCTGGCCC
238829

Qy         885  CCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCC 944
                | || ||| |||||||||| || ||||| ||||||||||||||||| ||||||||||||||
Db      238828  CTGAGGAAGTCAAGCAAAGTCCGGATGAGTTCTGCCACAGTGACTTCGAAGATGCCTCCC
238769

Qy         945  AAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGG 1004
                |||||||||||||||||||||||||||| || || ||||| ||||    ||||| |||||||
Db      238768  AAGGAAGCGACACAAGAATCTGCCCCCCCTCTTCGCTGCTTCCTCATGACCGAAAGTGGG
238709

Qy        1005  GTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATG 1064
                |||| |||||||||||||||||||| ||||||| || ||||||||||||||||| |||||||
Db      238708  GTTTTATTACTGTGGGTTACAGGGGATCCTGTACCTTGGGCAGAGAGGGGCAAGCAGATG
238649

Qy        1065  CCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAG 1124
                |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| ||||
Db      238648  CCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGCGGAAGGATTTCCTTGGCAAAGGAAG
238589
```

```
Qy        1125 TCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCT 1184
               |  ||||| |||||||||||||||||| |||||||| || ||||| |||||||||||||||
Db      238588 TTTTTGGTGAAACTTTGAATGAAAGTAGAGACCCCGACCGAGCTCCAGAAAGATACACCT
238529

Qy        1185 CCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTG 1244
               ||||||||||||||||| ||||| || |||||||| ||||||||||||||||||||||||
Db      238528 CCAGATTTTATCTCAAGTTCAAACATCTGGAAAGAGCTTTTGATATGTTGTCAGAGTGTG
238469

Qy        1245 GATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAG 1304
               |||||||||||||||||||||||||| ||||| |||||||||| |||||||| |||||||
Db      238468 GATTCCACATGGTGGCCTGTAACTCCTCGGTTACAGCATCTTTTGTCAACCAGTATACAG
238409

Qy        1305 ATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAG 1351
               |  |||||||||||||| ||||| ||||||||| ||||||||||| ||
Db      238408 AAGACAAGATCTGGTCGAGCTATACTGAATACGTCTTCTACCGTAAG 238362


RESULT 11
AC098707
LOCUS       AC098707              230128 bp    DNA      linear   ROD 21-JUN-2002
DEFINITION  Mus musculus clone RP23-1I13, complete sequence.
ACCESSION   AC098707
VERSION     AC098707.2  GI:19909459
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 230128)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230128)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 230128)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 230128)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Apr 3, 2002 this sequence version replaced gi:16554404.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
```

--.
FEATURES   Location/Qualifiers
      source   1. .230128
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10090"
                  /clone="RP23-1I13"
BASE COUNT   69218 a   47164 c   45440 g   68306 t
ORIGIN

  Query Match   23.4%;   Score 811.8;   DB 10;   Length 230128;
  Best Local Similarity 83.1%;   Pred. No. 6.8e-165;
  Matches 982; Conservative   0; Mismatches 172; Indels 27; Gaps   4;

Qy   187 ATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGAAGAT 246
              | || ||| ||||||||||||||| | ||||||| | | | |||| | |||||||||
Db   176288 AGACAGACCTCTGAGTAACTGGGCATTTGGCCTTCTTGCCTACAGAGCCTAATGGAAGAT
176347

Qy   247 TGGATATAGACGAGTTGATTAT---ATTTTATGAAGTAGCAGCTCACTACCATCCACCAT 303
              || || |||| ||| |||||| | | |||||||| ||||| ||| || |||
Db   176348 TGAATGTAGAGGAGCCGATTATTACATCCTGTGAAGTAAGAGCTCGCTATCAGCCA----
176403

Qy   304 CCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGATT 363
                | ||| |||| || | ||||||| | |||| |||||||||| || | | || ||
Db   176404 ----GCTTTTAACTTCTCCTCCAGCATTATTTCAACTGTGGACTCTCCTAAAATCTGCTT
176459

Qy   364 TCTTGGGGGAA-------------AAATACTGGGATAAGAGGAGGTCATTTTTTAATAAG 410
              | |||||| ||| ||| | ||| | ||| | || | | ||
Db   176460 TGTTGGGAGAATAAAAACCAAAACAAAACCAGGGTAAGAAGGGGTGATTTAAAAAGAAAA
176519

Qy   411 TTAGCATCCTTTTCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGG 470
              | |||| | | || || || || ||| | ||||| || |||||| |||| ||||
Db   176520 GTTAGTATCTTTATCTTACCTCACTCGTGGATGCAAAGTCTAGGGCTGTAACTCAGTTGG
176579

Qy   471 ATTGCACCTTTAAATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAA 530
              ||||||||||||| || | | || |||||||||||||||||||||||||| ||||| |
Db   176580 ATTGCACCTTTAATTCCA---GGAGGCTGCAGAAGAAAGGGACAATGGCTCTAAGTGGGA
176636

Qy   531 ACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGG 590
              ||||||| ||||||||||||| || || ||| | || |||||||||||||||||||||| |
Db   176637 ACTGTAGCCGTTATTATCCTCGGGACCAGGGGGCTGCTGTTCCCAACTCCTTCCCTGAAG
176696

Qy   591 TGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCA 650
              | |||||||||||||| || || || ||||| || |||||||||||||||||| |||||||
Db   176697 TCATAGAGCTGAATGTTGGCGGCCAGGTTTACTTCACTCGCCATTCCACATTAATAAGCA
176756

```
Qy         651 TCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAG 710
               |||| |||||| ||||||||||||||||||| |||||||||||||||||| ||||| |||||||
Db      176757 TCCCCCATTCTCTCCTGTGGAAAATGTTCTCCCCAAAGAGAGACACTGCTAACGATCTAG
        176816

Qy         711 CCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTC 770
               ||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Db      176817 CCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGCTTTCTGTTCCGTTATATTC
        176876

Qy         771 TGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGAC 830
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |
Db      176877 TGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAGAGGAAGGC
        176936

Qy         831 TGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATG 890
               ||||||| ||||||||| |||||||||||||| ||| | |||||||||||| ||||||| |
Db      176937 TGAAAAGAGAAGCTGAGTACTTCCAGCTCCCTGACCTCGTCAAACTCCTGGCCCCCGAGG
        176996

Qy         891 AAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAA 950
               | |||||||||||| |||| ||||| ||||||||||||||| |||||||||||||||||||||
Db      176997 ATGTCAAGCAAAGCCCGGATGAGTTCTGCCACAGTGACTTCGAAGATGCCTCCCAAGGAA
        177056

Qy         951 GCGACACAAGAATCTGCCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCA 1010
               ||||||| |||||||||||| || || |||||| ||| ||||||||||||||||||| |
Db      177057 GCGACACGAGAATCTGCCCCCCCTCTTCGCTGCTTCCTCACGACCGCAAGTGGGGTTTTA
        177116

Qy        1011 TTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGT 1070
               ||||||||||||||||| ||||||| ||||||||||||||||||| || |||||||||||||
Db      177117 TTACTGTGGGTTACAGGGGATCCTGTACCTTGGGCAGAGAGGGGCAAGCAGATGCCAAGT
        177176

Qy        1071 TTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTG 1130
               | |||||||| |||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db      177177 TCCGGAGAGTCCCCCGGATTTTGGTTTGCGGAAGAATTTCCTTGGCAAAAGAAGTCTTTG
        177236

Qy        1131 GAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGAT 1190
               ||||||||||||||||||||| ||||||| || ||||| |||||||||||||||||||||||
Db      177237 GAGAAACTTTGAATGAAAGTAGAGACCCCGACCGAGCTCCAGAAAGATACACCTCCAGAT
        177296

Qy        1191 TTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCC 1250
               ||||||||| || || || |||||||| ||||||||||||||||||||||||||||||||||
Db      177297 TTTATCTCAAGTTTAAACATCTGGAAAGAGCTTTTGATATGTTGTCAGAGTGTGGATTCC
        177356

Qy        1251 ACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACA 1310
               ||||||||||||||||||||| ||||| |||||||||||| ||||||| ||||||| || |
Db      177357 ACATGGTGGCCTGTAACTCCTCGGTTACAGCATCTTTTGTCAACCAGTATACAGAAGATA
        177416

Qy        1311 AGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAG 1351
               |||||||||| ||||| ||||| || |||||||| ||| ||
```

Db   177417 AGATCTGGTCGAGCTATACTGAGTACGTCTTCTATCGTAAG 177457

RESULT 12
AC112599
LOCUS   AC112599   249703 bp   DNA   linear   HTG 21-SEP-2002
DEFINITION   Rattus norvegicus clone CH230-112A20, *** SEQUENCING IN PROGRESS
                ***.
ACCESSION   AC112599
VERSION   AC112599.4 GI:23266003
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
REFERENCE   1 (bases 1 to 249703)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
                Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
                Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
                Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
                Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
                Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
                Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
                Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
                Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
                Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
                Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
                Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
                Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
                Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
                Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
                Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
                Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
                Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
                Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
                Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
                Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
                Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
                Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
                Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
                Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
                Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
                Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
                Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
                Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
                Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
                Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
                Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
                Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
                Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
                Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
                Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
                Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
                Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
                Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
                Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

          Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
          Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
          Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
          Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
          Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
          Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
          Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
          Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
          Weinstock,G. and Gibbs,R.A.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 249703)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 249703)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Sep 21, 2002 this sequence version replaced gi:21743383.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads. Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GRQH
                Center clone name: CH230-112A20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 233268 bases at least Q40
                Consensus quality: 235949 bases at least Q30
                Consensus quality: 237476 bases at least Q20
                Estimated insert size: 261159; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            ----
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced

```
                  * by the finished sequence as soon as it is available and
                  * the accession number will be preserved.
                  *       1   249703: contig of 249703 bp in length.
FEATURES                  Location/Qualifiers
     source             1. .249703
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-112A20"
     misc_feature       1. .1198
                        /note="wgs_end_extension
                        clone_end:T7"
     misc_feature       4045. .5985
                        /note="wgs_end_extension
                        clone_end:T7"
     misc_feature       complement(6997. .7899)
                        /note="clone_boundary
                        clone_end:T7
                        site:EcoRI
                        end_sequence:BH365326"
     misc_feature       138060. .196295
                        /note="clone_boundary
                        clone_end:Sp6
                        site:EcoRI
                        end_sequence:BH365327"
BASE COUNT     69486 a   47001 c   47902 g   74288 t   11026 others
ORIGIN

   Query Match            23.2%;   Score 806.2;  DB 2;  Length 249703;
   Best Local Similarity  69.2%;   Pred. No. 1.1e-163;
   Matches 1420;  Conservative   0;  Mismatches  573;  Indels   60;  Gaps   21;


Qy      1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
             ||||||| ||||| ||||| |||| || || ||||||||||||||||||||||||| || |
Db      159149 GTGAGCCCTCCAGGTGGTCCTCCTCCCATTGTGATTGCTGCTGCAAGAATGGCAAGGGAG
159208


Qy      1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
             |||    ||||||||||| ||||| || ||||||||||||||| || || |||||||||||
Db      159209 ACA---AAGGGGAGAGTGGCACTTCCTGCAATGACCTCTCTACTTCCAGCTGCGACAGCC
159265


Qy      1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| |||||||||||||||||||||||| || ||||||||| || || || || ||| ||
Db      159266 AGTCAGAGGCCAGCTCTCCCCAGGAGACAGTGATCTGTGGGCCTGTAACGCGTCAGGGCA
159325


Qy      1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
             ||||||||||||||||||| ||||||||||| |||| || ||||||||||||||||||||
Db      159326 ACATCCAGACTCTGGACCGGCCCATCAAGAAAGGCCCCGTGCAGCTGATCCAACAGTCAG
159385


Qy      1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
             ||||| |||||||||| ||| | |||||||| |||||||| ||||| ||||| |||||||
Db      159386 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACTTCCGGCTCTAGGGAGTCGAACA
159445
```

```
Qy       1647 TGAGCAGCAAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
              | ||||||||||| ||||||| || ||||||| |||| || |||||| ||||||||||||||
Db     159446 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGGTCTCCATCGAGGAAGAGCTGGAGAAAT
       159505

Qy       1707 GTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
              ||||||||||||||||| || | |||||||||||||| || |||||||||||||||||||||
Db     159506 GTATCCAGGATTTCCTGAAGATAAAAATTCCAGATCGCTTCCCTGAGAGAAAACATCCTT
       159565

Qy       1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
              |||| |||||||||||| |||||||||||||||| ||||||||| | ||| | |
Db     159566 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGAGGGCAGTGGGTAGTCA----- 
       159620

Qy       1827 AAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
                        || ||||||||| |||||| | ||||||| ||||| |||||||||||||
Db     159621 ----------CCACTTTGAAATAAACCTCCTGAAAGGAAGACATATATTAAAGGAAAAA
       159669

Qy       1887 AATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTA 1946
              | |||||||| || || || | |||||||||||| |||||||||| | ||||||||||
Db     159670 TA-ACAACTAACAATCCATATGTGTTAGAACACAAT-GTCCATTGATGTCCTACTGCCTA
       159727

Qy       1947 CTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA 2006
              |||| |||||| ||||||||||||||||| |||||||||||||||||||||||||||||||
Db     159728 CTTTGCCTAGCTCACCTTAACATGTAAATTCACAGGGTAGATTTCTTTCTAGATGTGGAA
       159787

Qy       2007 GTACAAGAAAATCTTTTTTAGTTATTTGT-----TTGTTTACTTCGTCCCATGTGCTAAC 2061
              | ||| | | ||| | | | || ||||||| ||||||||| | |
Db     159788 CCAGAAGCGATGCCCTTATGCTGTCCTCTGTCTCTTATTTACTTGGTCCCATGTGTTGAG
       159847

Qy       2062 TATCTT-ATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCAT 2120
              ||||| |   |  ||  ||||  ||||||||| |          |  |        |  | |
Db     159848 AATCTTAAGGTTCAAGGAGAACCAGCTACGTGAGTAGCTCGAATCCCAAACCTGCTTTTT
       159907

Qy       2121 TTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGGGTATG 2180
              | |     |  |||| |||| |  | || | ||| || || | |     | || | |
Db     159908 TGTTTGTTTGTTTTGTTTCCTCTCATTTTCTGCCTCCTTCC-CTTGACCAAGAATGGACA
       159966

Qy       2181 GTAGAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTTCCTTT 2240
              || ||| |||| | |||| |||||||| ||| ||| ||| || | |||| ||| | |||
Db     159967 GTTGAAGGAGATATAACCCAGTGGCATATGTTAAGAAATTATTCTTTTTCCTTTACTTTT
       160026

Qy       2241 TGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTATATGTG 2300
              ||     |||     |||| |||  |||||||||||| |||| ||| |||| ||||||||||
Db     160027 GTTTATGGGGTGAGGGGAGAACGGCAGATTTGTATGATTTTCCACTAAAATCTCTATGTG
       160086

Qy       2301 CCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAACACAAGCA-CAACTAAGTAT 2359
              |||| || ||||||||| |||||||||| ||| || ||| || | |  | | |   || |
```

```
Db    160087  CCAGGTTCTATTGACTTTGTATGCATGAGCGTTTCTGACACAAGCACAGTATATGTCTGT
      160146

Qy      2360  GTATATACACATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTCCTGC  2419
              ||||| ||||          |||     ||  || || |    |   |  ||||||     |||
Db    160147  ATATATGCACAAAGAATGCACATGACCCAGGGCTGGGACAGCAGAGGGCTAACACCTTAC
      160206

Qy      2420  TCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTAAAAT  2479
              | ||||| |||| ||  |||       ||||||        ||| |||| || ||    |||||
Db    160207  TGCCAGCTGCCC-CTTCAAGAGCGCTTCAGACAACAAAGCCTCTGTCTATTCAGTAAAAC
      160265

Qy      2480  TCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTCT-AACATCAAATGACTCTC  2538
              ||  |||| ||||||| |||||||   |||  |||||||||||| ||  |  |  | |  |  |
Db    160266  CCTCCTGGGCAGATTTGCCAGCCTCCCTTGGCAACACTTTCTAAAGCTGTATAGGCCCCC
      160325

Qy      2539  ATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTACTGC  2598
              ||||||||||||    | |||| || ||||| |||||||| ||||||||| ||       ||
Db    160326  ATCATCAACAAAT----TCCCTTTTTTTGAAACAAATACCCGCAGGCTCCTTTGATTTAC
      160381

Qy      2599  TTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGTCAAACATTCCTTGTGTTA  2658
              | |||||||| || ||| | ||| |    ||||  |||| ||| |||   |  |||||||
Db    160382  TGTGCTCTTTCCCTACATCAGGAGCTTGTGAGATGAGCTAGTCTAACCCTGTTTGTGTTT
      160441

Qy      2659  AA---AAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATGACTCCACACTCAGCC  2715
              ||    | || |||||| |||  ||||||||||     || | ||   ||  ||||||||||
Db    160442  AACAGACAAGCAAACAGTCACATCCACAAACAGAGCTTC-AAGACACCACCTACTCAGCC
      160500

Qy      2716  TTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCAGAGTTCTTACTGCCC  2775
              |||| |   || |    |||| |||||| || |||    || |       ||      |||||
Db    160501  TTCTCCATTCTTACTAGAATGATCACCATTCTCTAGCTGACTCAGAGTTTTAACTTGCCC
      160560

Qy      2776  ACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCTGTGTTCCTGTGTTGT  2835
              ||| |||| |        | ||||||||| | |||||||||   || |   |   | |     |
Db    160561  ACATTTTATTAAAGAGGCCTTT-GATATAATCCAGGCAAATTCTTTGCATACCTGTGGTT
      160619

Qy      2836  TGTGGAACTAAGGACAACACACA-----GTACTTGAATAAGGGTCCGGCCTTTTGTTTGT  2890
              |||| | | | | ||| |  || ||     |  |   | |   |   |   ||| |     |
Db    160620  TGTGAAGCAATGAACTAATTAAACATGCATCCAGCCTTCTGTTCTCTTGTTTTAGAGGAT
      160679

Qy      2891  TTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAATTTTAAACTAC  2950
              || |          | ||  |    || |       || | |||||||| ||| ||| |||   |||||
Db    160680  TTGTGTCCCCCCCCTCCCCCGCCACATACATCTTAATTTCTCATACAAACTTTCCACTAC
      160739

Qy      2951  AAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATTC-GGGCTTTG  3009
              |     |||| | || |||| |||| ||    |||| | ||||||||| || | |||  ||||
Db    160740  ACCTATACACTGTTGTTTGCCTGTATCCAGGTTTGGATACCTTTGGAATCCTGGGGTTTG
      160799
```

```
Qy        3010 GCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACAACCCAACAA 3069
               ||||| || | ||| | ||||| ||| ||||||| ||    ||||||||| || ||||
Db      160800 ACTGTGGCCCTACTATGGTTTAGTTGTATCATTTCTACAGTGTCTGTAATAATCCAAGTG
160859


Qy        3070 GGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCTTTTTCAATT 3129
               ||| |||| || |   | ||| ||    || |||  |     | ||||||
Db      160860 GGTGACTGGAACATAAAGGTTTCTAATTTGATTTTTTA--------AACTTTTTTTAA
160911


Qy        3130 ACATCCTGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCCGTCCATGTA 3189
               ||||||| |||||||| ||   | ||||||||| ||| || | |  || |||||||
Db      160912 TAGTCCTGACCTGTATAGATACCATCCAAAAGAAATTGTGAACA-CTGTCTATCCATGTG
160970


Qy        3190 ACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTA 3249
               ||||||| | || ||| ||||| ||| ||||  |  | |||| || ||||| ||
Db      160971 ACTCTGTCATCTATTAATCTACCAGTAGTTCTTCTGTTCACCTGTGTTAAAAGAATGTCC
161030


Qy        3250 AAAGACCAGTTTT-ATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATT 3308
               |||||| | |||| |||||||||||||||||||||||| || |||||||||||  ||||| |
Db      161031 AAAGACAACTTTTAATTTTCAGCATTCCTCATATATCTCAGTGGTAACTGAAAAAGACGA
161090


Qy        3309 TGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATG 3368
               | |     | |     ||||||||||||||||| |||| |||| || | ||| |||||||| ||
Db      161091 TTT---ATCACTAGTGTGTGCCAAGCATTCCT-ATTTTTGTTTGTGTGTGTGTGTGTG
161146


Qy        3369 TGTGTATGTGTAT 3381
               ||||| ||||| |
Db      161147 TGTGTGTGTGTGT 161159



RESULT 13
BC049734
LOCUS       BC049734                781 bp    mRNA     linear   ROD 01-APR-2003
DEFINITION  Mus musculus, clone IMAGE:6771233, mRNA.
ACCESSION   BC049734
VERSION     BC049734.1  GI:29436685
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 46 Row: g Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

```
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6771233"
                     /tissue_type="Testicle, mouse"
                     /clone_lib="NIH_MGC_169"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
BASE COUNT       348 a    146 c    169 g    118 t
ORIGIN
```

```
  Query Match               10.6%;   Score 367.4;  DB 10;   Length 781;
  Best Local Similarity     80.2%;   Pred. No. 8.5e-69;
  Matches  463; Conservative    0; Mismatches   96;  Indels    18;  Gaps     2;


Qy      1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
             ||||||||||| | ||||| |||| || || || ||||||||||||||||||||| || |
Db       101 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG 160

Qy      1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
             |||    |||| |||||||||| || ||||||||||| ||||| || ||||| |||||||
Db       161 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC 217

Qy      1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| ||||||||||||||| |||||||||| ||||||||| |||||| || || || ||
Db       218 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 277

Qy      1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
             ||||||||||||||||| || |||||||||| || || || |||||||||||||||||||
Db       278 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 337

Qy      1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
             ||||| |||||||||| ||| | ||||||| |||||| |||||||||||||| |||||||
Db       338 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA 397
```

```
Qy   1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
           | ||||||||||| ||||||| || ||||||||||| || ||||| |||||||||||||
Db    398 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT 457

Qy   1707 GTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
           ||||||||||||||| | || | ||||||||||||||||| || ||||||| ||||||||||||
Db    458 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAAACATCCTT 517

Qy   1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
           |||| ||||||||||||| |||||||||||||||||||| ||| || | |||
Db    518 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGGAGGGCTGTGG---------- 566

Qy   1827 AAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
            |    | | || ||||||||| |||||| ||||||| ||||| |||||||||||||
Db    567 ----GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAGACATATGTTAAAGGAAAAA 622

Qy   1887 AATACAACTAATGATGCACATTTCTTAGAACACAATA 1923
           || | || || | || | || | || | ||
Db    623 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 659
```

RESULT 14
BX323465

```
LOCUS       BX323465            175059 bp    DNA     linear   HTG 06-JUN-2003
DEFINITION  Danio rerio clone DKEYP-9C6, *** SEQUENCING IN PROGRESS ***, 3
            unordered pieces.
ACCESSION   BX323465
VERSION     BX323465.4  GI:31559295
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio .
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 175059)
  AUTHORS   Mclaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 9, 2003 this sequence version replaced gi:29825526.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp9C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 174665 bases at least Q40
            Consensus quality: 174734 bases at least Q30
            Consensus quality: 174818 bases at least Q20
            Insert size: 174859; sum-of-contigs
            Insert size: 176642; 1.7% error; agarose-fp
```

Quality coverage: 10.25x in Q20 bases; sum-of-contigs Quality
coverage: 10.15x in Q20 bases; agarose-fp
---------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1   44665: contig of 44665 bp in length
*   44666   44765: gap of 100 bp
*   44766   172724: contig of 127959 bp in length
*   172725   172824: gap of 100 bp
*   172825   175059: contig of 2235 bp in length.
FEATURES   Location/Qualifiers
     source   1. .175059
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-9C6"
                     /clone_lib="DanioKeypilot"
     misc_feature   1. .44665
                     /note="assembly_fragment:01355
                     clone_end:T7
                     vector_side:left"
     misc_feature   44766. .172724
                     /note="assembly_fragment:00935.0"
     misc_feature   172825. .175059
                     /note="assembly_fragment:01078"
BASE COUNT   56742 a 30793 c 31197 g 56126 t   201 others
ORIGIN

   Query Match   9.2%;   Score 319.2;   DB 2;   Length 175059;
   Best Local Similarity   63.6%;   Pred. No. 2.3e-58;
   Matches   533; Conservative   0; Mismatches 278; Indels   27; Gaps   2;

Qy   514 AATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTCC 573
              |||||| |||| || ||| || || |||| | | | | | | | ||
Db   127701 AATGGCCCTGACTGAAAATTGCAGGACTTATCAAACGCCCAAGGACAGTGGATGTGCTCA
127760

Qy   574 CAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGGTCAAGTTTATTTTACTCGCCA 633
               | | || |||| ||||| ||||| ||||| ||||| || || || || | |||||||||
Db   127761 GAGTTGCTCTTCTGATGTGGTGGAGCTCAATGTAGGTGGACAGGTGTACTACACTCGCCA
127820

Qy   634 TTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGA 693
               | |||| | | || | || | || || ||| | | ||| |||| ||| | || | |||
Db   127821 TGCCACCCTCACCAGTGTGCCAAACTCACTGCTGGGTAAATTGTTCTCCTCTAAAAAGA
127880

Qy   694 CACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGATT 753
               || |||| || || | ||||| ||||||| | | ||||| ||| | || |||||
Db   127881 CATTTCTAACGACCTCACGCAGGACATCAAGGGACGCTACTTCATCGACCGGGACGGATT
127940

```
Qy        754 CTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTT 813
              | ||  | ||  |  ||||||||||||| | || | |    ||  |||||||| ||| | ||
Db     127941 TCTCTTTAGGTACGTGCTGGACTATCTCCGCGATAAGACTGTCGTCCTGCCGGATTATTT
       128000

Qy        814 TCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCAA 873
              ||| || || || || || ||||||| | ||||||||| |  ||||||||| || ||  | |||||
Db     128001 TCCGGAGAAGGGGAGGCTGAAACGCGAAGCTGAGTTTTTCCAGCTGCCCGAGCTCGTCAA
       128060

Qy        874 ACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTGA 933
              | ||||                    || |||||||||| |   | |||||||||| |||||
Db     128061 AATCCT--------------------AAACCCAGATGATTATAGTCACAGTGATTTTGA
       128099

Qy        934 AGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGA 993
              || || ||||| ||||||||| || |  | | | | ||  | ||| | ||||  | |   |||
Db     128100 CGAAGCATCCCAGGGAAGCGACCAGAGGTTATATCCAGCCTCTTACCTGGACGCGCGCGA
       128159

Qy        994 CCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGG 1053
              |||       |  || ||||| || || || ||||   | || ||| | || || || ||
Db     128160 CCGACGCTACGGCTTCATCACGGTCGGGTACAAGAGCTCGTGCGCATTCGGGAGGGACA-
       128218

Qy       1054 ACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTT 1113
               | ||| ||||    || || | |||  || || | |||   || || |||| ||||| |    |
Db     128219 -----CTGATCCCAAAGCCCGCGGAATACCCAAAATCTTCATTTGCGGAAGAGTCGGTCT
       128273

Qy       1114 GGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGA 1173
              ||| |||||||| |||| || || || | | || || ||||| || |||||  || | || ||
Db     128274 GGCGAAAGAAGTTTTTCGGCGACGCACTAAACGAGAGCAGGGATCCTGACAGACCGCCGGA
       128333

Qy       1174 AAGATACACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTT 1233
              | ||||| ||       ||||| || || || || ||||||||   | || ||||||||| |
Db     128334 GCGTTACACTTCTCAGTTTTACCTGAAGTTTCGCCACCTGGAGCGAGCGTTTGATATGCT
       128393

Qy       1234 GTCAGAGTGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTTCATCAA 1293
              | ||| | || ||||||||| || || || || || ||||| | || ||||      |||
Db     128394 CGCGGAGAGCGGGTTCCACATCGTCGCGTGCAATTCATCACTCACCACATCTCCTCACAA
       128453

Qy       1294 CCAATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAG 1351
              |   || | |||||| |      |||||| | | ||| || || |||||||| ||| ||
Db     128454 CAGGCATGCTGATGATAGATACTGGTCCAACAACACAGAGTACGTCTTCTATCGTAAG 128511


RESULT 15
BX470157
LOCUS        BX470157              200467 bp    DNA      linear    HTG  05-MAY-2003
DEFINITION   Danio rerio clone CH211-119P14, *** SEQUENCING IN PROGRESS ***, 8
             unordered pieces.
ACCESSION    BX470157
```

```
VERSION       BX470157.2  GI:30387082
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE        Danio rerio (zebrafish)
  ORGANISM    Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE     1  (bases 1 to 200467)
  AUTHORS     McLay,K.
  TITLE       Direct Submission
  JOURNAL     Submitted (04-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT       On May 5, 2003 this sequence version replaced gi:30349786.
              -------------- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-help@sanger.ac.uk
              -------------- Project Information
              Center project name: zC119P14
              -------------- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Chemistry: Dye-terminator; 100% of reads
              Consensus quality: 198546 bases at least Q40
              Consensus quality: 199010 bases at least Q30
              Consensus quality: 199314 bases at least Q20
              Insert size: 199767; sum-of-contigs
              Insert size: 201190; 3.3% error; agarose-fp
              Quality coverage: 5.95x in Q20 bases; sum-of-contigs Quality
              coverage: 6.07x in Q20 bases; agarose-fp
              --------------.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 8 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *          1      9514: contig of 9514 bp in length
              *       9515      9614: gap of 100 bp
              *       9615     14582: contig of 4968 bp in length
              *      14583     14682: gap of 100 bp
              *      14683     18933: contig of 4251 bp in length
              *      18934     19033: gap of 100 bp
              *      19034     66645: contig of 47612 bp in length
              *      66646     66745: gap of 100 bp
              *      66746     73558: contig of 6813 bp in length
              *      73559     73658: gap of 100 bp
              *      73659    113761: contig of 40103 bp in length
              *     113762    113861: gap of 100 bp
              *     113862    193754: contig of 79893 bp in length
              *     193755    193854: gap of 100 bp
              *     193855    200467: contig of 6613 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200467
```

```
                        /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /clone="CH211-119P14"
                        /clone_lib="CHORI-211"
        misc_feature    1. .9514
                        /note="assembly_fragment:01971
                        fragment_chain:1
                        clone_end:SP6
                        vector_side:left"
        misc_feature    9615. .14582
                        /note="assembly_fragment:01296
                        fragment_chain:1"
        misc_feature    14683. .18933
                        /note="assembly_fragment:00946
                        fragment_chain:1"
        misc_feature    19034. .66645
                        /note="assembly_fragment:00199
                        fragment_chain:1"
        misc_feature    66746. .73558
                        /note="assembly_fragment:02277
                        fragment_chain:2"
        misc_feature    73659. .113761
                        /note="assembly_fragment:01029
                        fragment_chain:2"
        misc_feature    113862. .193754
                        /note="assembly_fragment:00416
                        fragment_chain:2"
        misc_feature    193855. .200467
                        /note="assembly_fragment:02202
                        fragment_chain:2
                        clone_end:T7
                        vector_side:right"
BASE COUNT      64222 a   35134 c   35671 g   64740 t      700 others
ORIGIN
```

   Query Match   9.2%;   Score 317.6;   DB 2;   Length 200467;
   Best Local Similarity   63.5%;   Pred. No. 5.2e-58;
   Matches 532; Conservative   0; Mismatches 279; Indels   27; Gaps   2;

```
Qy         514 AATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTCC 573
               ||||||  ||||  ||  |||  ||  ||    ||||  |   | |    | |  |       | ||
Db      147426 AATGGCCCTGACTGAAAATTGCAGGACTTATCAAACGCCCAAGGACAGTGGATGTGCTCA
147485

Qy         574 CAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCCA 633
                |  |  ||     ||||  |||||  |||||  |||||  ||  ||  ||  ||  ||  |  |||||||||
Db      147486 GAGTTGCTCTTCTGATGTGGTGGAGCTCAATGTAGGTGGACAGGTGTACTACACTCGCCA
147545

Qy         634 TTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGA 693
                |   |||  | |   || |  || |  | ||  || |||  | |||  ||||  |||  | ||  |  |||
Db      147546 TGTCACCCTCACCAGTGTGCCAAACTCACTGCTGGGTAAATTGTTCTCCTCTAAAAAAGA
147605

Qy         694 CACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGATT 753
                ||     ||||  ||  ||    |  |||||   |||||||  | |   |||||  |||  |  ||  |||||
```

```
Db     147606 CATTTCTAACGACCTCACGCAGGACATCAAGGGACGCTACTTCATCGACCGGGACGGATT
147665

Qy        754 CTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTT 813
              I II  I II  I IIIIIIIIIIIII I II I I   II IIIIIIII III I II
Db     147666 TCTCTTTAGGTACGTGCTGGACTATCTCCGCGATAAGACTGTCGTCCTGCCGGATTATTT
147725

Qy        814 TCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCAA 873
              III II II II II IIIIII I IIIIIIII I  IIIIIIII II II I IIIII
Db     147726 TCCGGAGAAGGGGAGGCTGAAACGCGAAGCTGAGTTTTTCCAGCTGCCCGAGCTCGTCAA
147785

Qy        874 ACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTGA 933
              I IIII               II IIIIIIIII I   I IIIIIIIII IIIII
Db     147786 AATCCT-------------------AACCCCAGATGATTATAGTCACAGTGATTTTGA
147824

Qy        934 AGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGA 993
              II II IIIII IIIIIIIII  II I I II I II I IIII I I   III
Db     147825 CGAAGCATCCCAGGGAAGCGACCAGAGGTTATATCCAGCCTCTTACCTGGACGCGCGCGA
147884

Qy        994 CCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGG 1053
              III    I  II IIIIII II II II IIII  I II III I II II II II
Db     147885 CCGACGCTACGGCTTCATCACGGTCGGGTACAAGAGCTCGTGCGCATTCGGGAGGGACA-
147943

Qy       1054 ACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTT 1113
               I III IIII      II  II I III  II II IIII IIIII  I        I
Db     147944 -----CTGATCCCAAAGCCCGCGGAATACCCAAAATCTTCATTTGCGGAAGAGTCGGTCT
147998

Qy       1114 GGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGA 1173
              III IIIIIIII II II II  I  I II II III I II IIIII  II I II II
Db     147999 GGCGAAAGAAGTTTTCGGCGACGCACTAAACGAGAGCCGGGATCCTGACAGACCGCCGGA
148058

Qy       1174 AAGATACACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTT 1233
              I IIIIII II     IIIII II II II    IIIIIIII I II IIIIIIIIII I
Db     148059 GCGTTACACTTCTCAGTTTTACCTGAAGTTTCGCCACCTGGAGCGAGCGTTTGATATGCT
148118

Qy       1234 GTCAGAGTGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTTCATCAA 1293
              I III I II IIIIIIIII II II II II IIIII  I II  IIIII I  III
Db     148119 CGCGGAGAGCGGGTTCCACATCGTCGCGTGCAATTCATCACTCACCACATCTCCCCACAA
148178

Qy       1294 CCAATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAG 1351
              I     II I IIIIII I   IIIIII I I IIII II II IIIIIIIII III II
Db     148179 CAGGCATGCTGATGATAGATACTGGTCCAACAACACAGAGTACGTCTTCTATCGTAAG 148236
```

Search completed: January 29, 2004, 02:30:04
Job time : 12473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 20:04:04 ; Search time 867 Seconds
                                             (without alignments)
                                             10797.752 Million cell updates/sec

Title: US-10-056-884A-1
Perfect score: 3468
Sequence: 1 caagcactgtgctaaagtgt..........aaaaaaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_19Jun03:*
                 1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
                 25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

        Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3468	100.0	3468	24	AAD46068	Human K+betaM2 cDN
	2	1640.8	47.3	2412	24	ABN59764	Novel human coding
	3	769	22.2	769	24	AAD46125	Human BAC AC008652
c	4	699.2	20.2	906	22	ABA09216	Human VM106R.1 hom
	5	423.4	12.2	440	22	AAS34230	Human cDNA encodin
	6	319.4	9.2	2398	25	AAD49513	Human TRICH-15 cDN
	7	205	5.9	632	24	ABV99059	Human pancreatic c
	8	201	5.8	614	24	ABV95156	Human pancreatic c
	9	167	4.8	2052	24	ABT09812	Polynucleotide enc
c	10	167	4.8	109201	24	ABQ88125	Human osteoblast d
	11	114.2	3.3	854	24	ABQ40654	Oligonucleotide fo
c	12	114.2	3.3	854	24	ABQ40655	Oligonucleotide fo
	13	109.8	3.2	1757	24	ABQ13668	Oligonucleotide fo
c	14	109.8	3.2	1757	24	ABQ13669	Oligonucleotide fo
c	15	108.6	3.1	854	24	ABQ40656	Oligonucleotide fo
	16	108.6	3.1	854	24	ABQ40657	Oligonucleotide fo
	17	104.6	3.0	688	24	ABT09813	K+beta M6 related
c	18	95.8	2.8	1757	24	ABQ13666	Oligonucleotide fo
	19	95.8	2.8	1757	24	ABQ13667	Oligonucleotide fo
c	20	80	2.3	80	24	AAD46069	Antisense oligonuc
c	21	79	2.3	425	22	AAS60450	Human cancer agent
	22	79	2.3	1119	21	AAC60033	Human secreted pro
	23	79	2.3	1492	21	AAC98102	Human colon cancer
	24	79	2.3	1493	22	AAH34433	Human colon cancer
	25	79	2.3	1493	24	ABL90331	Human polynucleoti
c	26	76.8	2.2	2796	24	ABL90605	Human polynucleoti
	27	76.2	2.2	847	23	ABL06735	Drosophila melanog
c	28	76.2	2.2	2847	23	ABL06734	Drosophila melanog
	29	75.2	2.2	1856	23	ABK43528	DNA encoding novel
c	30	75	2.2	442	24	ABL94107	Arabidopsis thalia
c	31	74.2	2.1	655	22	AAH70113	Human cervical can
	32	74.2	2.1	887	21	AAC59297	Human secreted pro
	33	74	2.1	664	21	AAA26336	Human secreted pro
	34	73.8	2.1	1091	22	AAC89723	Maize ZmGnsN1-1 gl
	35	73.8	2.1	1091	25	ABX95035	cDNA encoding maiz
	36	73.8	2.1	1992	22	AAF72748	Human prostate can
	37	73.6	2.1	1204	21	AAC59836	Human secreted pro
c	38	73.4	2.1	375	23	ABV44911	Human prostate exp
	39	73.4	2.1	2440	22	AAH34932	Human colon cancer
c	40	73.4	2.1	4055	22	AAI58815	Human polynucleoti
	41	73.2	2.1	2377	21	AAC96941	Human secreted pro
c	42	73	2.1	348	22	AAL10133	Human breast cance
	43	73	2.1	1814	25	ABT17358	Human SLC7 related
	44	72.6	2.1	346	23	ABV48988	Human prostate exp
	45	72.2	2.1	297	22	AAS29114	cDNA encoding for

ALIGNMENTS

RESULT 1
AAD46068
ID   AAD46068 standard; cDNA; 3468 BP.
XX
AC   AAD46068;
XX
DT   27-DEC-2002 (first entry)
XX
DE   Human K+betaM2 cDNA.
XX
KW   Human; potassium channel beta-subunit; K+betaM2 protein; neural disorder;
KW   reproductive disorder; metabolic disorder; premature puberty; nephritis;
KW   endocrine disorder; memory disorder; neuroendocrine condition; asthma;
KW   spermatogenesis; renal disease; learning deficiency; Alzheimer's disease;
KW   neurodegenerative disease; proliferative disorder; autoimmune disease;
KW   carcinoid tumour; blood coagulation disease; blood platelet disease;
KW   rheumatoid arthritis; allergy; hyperproliferative disease; gene therapy;
KW   graft-versus-host disease; organ rejection; antisterility; thrombolytic;
KW   antiinflammatory; neuroprotective; anti-Parkinsonian; immunosuppressive;
KW   nephrotropic; cytostatic; nootropic; hypotensive; vulnerary; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   515..1801
FT   /*tag= a
FT   /product= "Human K+betaM2 protein"
XX
PN   WO200266601-A2.
XX
PD   29-AUG-2002.
XX
PF   24-JAN-2002; 2002WO-US02332.
XX
PR   24-JAN-2001; 2001US-263872P.
PR   14-FEB-2001; 2001US-269794P.
XX
PA   (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI   Feder J, Lee L, Chen J, Jackson D, Ramanathan C, Siemers N;
PI   Chang H, Carroll P;
XX
DR   WPI; 2002-691617/74.
DR   P-PSDB; AAE28618.
XX
PT   New potassium channel beta-subunit, K+betaM2, proteins and nucleic
PT   acids, useful for diagnosing, treating and/or preventing e.g.
PT   reproductive, neural, metabolic, endocrine, memory, neurodegenerative
PT   disorders or diseases -
XX
PS   Claim 1; Page 344-347; 366pp; English.
XX
CC   The present invention relates to human potassium channel beta-subunit
CC   (K+betaM2) proteins and polynucleotides encoding such proteins. The
CC   K+betaM2 sequences are useful for diagnosing, treating and/or preventing
CC   reproductive disorders, neural disorders, disorders related to aberrant

CC   potassium regulation or hyper potassium channel activity, metabolic
CC   disorders (e.g. premature puberty), endocrine disorders (e.g. aberrant
CC   growth hormone synthesis and/or secretion), memory disorder, disorders
CC   of the testis (e.g. spermatogenesis), neuroendocrine condition related
CC   to aberrant thyroid hormone release, renal disease or disorders (e.g.
CC   nephritis), disorders related to aberrant higher brain function (e.g.
CC   learning deficiencies), neurodegenerative diseases (e.g. Alzheimer's
CC   disease), proliferative disorders (e.g. carcinoid tumour) and disorders
CC   involving excessive smooth muscle tone or excitability (e.g. asthma).
CC   They may be used to modulate haemostatic or thrombolytic activity, to
CC   treat or prevent blood coagulation diseases or disorders, blood platelet
CC   diseases, wounds, autoimmune diseases, disorders or conditions (e.g.
CC   rheumatoid arthritis), allergic reactions (e.g. asthma), organ rejection
CC   or graft-versus-host disease, and hyperproliferative diseases. K+betaM2
CC   sequences are also used in gene therapy. The present sequence is human
CC   K+betaM2 cDNA.
XX
SQ   Sequence 3468 BP; 1038 A; 728 C; 703 G; 999 T; 0 other;

    Query Match   100.0%; Score 3468; DB 24; Length 3468;
    Best Local Similarity   100.0%; Pred. No. 0;
    Matches 3468; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CAAGCACTGTGCTAAAGTGTTTTTCATATGTCATGAAAAGTTGTGCCAGAAAATTATGGT 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CAAGCACTGTGCTAAAGTGTTTTTCATATGTCATGAAAAGTTGTGCCAGAAAATTATGGT 60

Qy   61 TTGAACATGGGCAGTTTTCTCCTACCGTCAGCTATATCCACAAGCATCACATGAAGTGGA 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTGAACATGGGCAGTTTTCTCCTACCGTCAGCTATATCCACAAGCATCACATGAAGTGGA 120

Qy   121 GATCTGGCAGCTCTGTGTATTTCAGTCAAGTTCCACAATGAAACCTGACAATAATGGTAA 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GATCTGGCAGCTCTGTGTATTTCAGTCAAGTTCCACAATGAAACCTGACAATAATGGTAA 180

Qy   181 AAACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATG 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATG 240

Qy   241 GAAGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCAC 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAAGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCAC 300

Qy   301 CATCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTG 360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CATCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTG 360

Qy   361 ATTTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCT 420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ATTTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCT 420

Qy   421 TTTCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTT 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TTTCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTT 480

Qy   481 TAAATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCG 540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    481 TAAATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCG 540

Qy    541 TTATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCT 600

Qy    601 GAATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTC 660

Qy    661 CCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTC 720

Qy    721 CAAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCT 780

Qy    781 CAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGA 840

Qy    841 AGCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCA 900

Qy    901 AAGCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAGCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAG 960

Qy    961 AATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGG 1020

Qy   1021 TTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGT 1080

Qy   1081 TCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTT 1140

Qy   1141 GAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAA 1200

Qy   1201 ATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGC 1260

Qy   1261 CTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTC 1320

Qy   1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGA 1380
```

```
Qy    1381 TTGCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTGCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGA 1440

Qy    1441 CCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCAT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCAT 1500

Qy    1501 CTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGG 1560

Qy    1561 CCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCT 1620

Qy    1621 GACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCT 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCT 1680

Qy    1681 CTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGA 1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGA 1740

Qy    1741 TCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATA 1800

Qy    1801 AGGGAGGGCTGGGGGCGGGGAAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAA 1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 AGGGAGGGCTGGGGGCGGGGAAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAA 1860

Qy    1861 AAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACA 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 AAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACA 1920

Qy    1921 ATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACA 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 ATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACA 1980

Qy    1981 GGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 GGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTTTTTTAGTTATTTGTTTGTT 2040

Qy    2041 TACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGC 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGC 2100

Qy    2101 TGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCC 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCC 2160

Qy    2161 CTCCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CTCCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTT 2220
```

```
Qy    2221 TTAATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTT 2280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 TTAATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTT 2280

Qy    2281 TTCACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAA 2340
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 TTCACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAA 2340

Qy    2341 CACAAGCACAACTAAGTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTC 2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CACAAGCACAACTAAGTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTC 2400

Qy    2401 CCAAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCT 2460
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 CCAAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCT 2460

Qy    2461 TCTGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTC 2520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 TCTGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTC 2520

Qy    2521 TAACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCT 2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TAACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCT 2580

Qy    2581 CAGGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGT 2640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CAGGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGT 2640

Qy    2641 CAAACATTCCTTGTGTTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATG 2700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 CAAACATTCCTTGTGTTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATG 2700

Qy    2701 ACTCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCA 2760
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 ACTCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCA 2760

Qy    2761 GAGTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCT 2820
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 GAGTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCT 2820

Qy    2821 GTGTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCC 2880
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 GTGTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCC 2880

Qy    2881 TTTTGTTTGTTTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAAT 2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 TTTTGTTTGTTTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAAT 2940

Qy    2941 TTTAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATT 3000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 TTTAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATT 3000

Qy    3001 CGGGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACA 3060
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 CGGGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACA 3060

Qy    3061 ACCCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCT 3120
```

```
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061   ACCCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAACTATCT 3120

Qy    3121   TTTTCAATTACATCCTGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCC 3180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3121   TTTTCAATTACATCCTGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCC 3180

Qy    3181   GTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACA 3240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3181   GTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACA 3240

Qy    3241   AGAAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAA 3300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3241   AGAAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAA 3300

Qy    3301   AAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTG 3360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3301   AAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTG 3360

Qy    3361   TGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAA 3420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3361   TGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAA 3420

Qy    3421   AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db    3421   AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468


RESULT 2
ABN59764
ID    ABN59764 standard; cDNA; 2412 BP.
XX
AC    ABN59764;
XX
DT    28-JUN-2002   (first entry)
XX
DE    Novel human coding sequence SEQ ID NO: 175.
XX
KW    Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW    antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW    neuroprotective; antiparkinsonian; protein therapy; EST;
KW    expressed sequence tag; gene; ss.
XX
OS    Homo sapiens.
XX
PN    WO200222660-A2.
XX
PD    21-MAR-2002.
XX
PF    10-SEP-2001; 2001WO-US26015.
XX
PR    11-SEP-2000; 2000US-0659671.
XX
PA    (HYSE-) HYSEQ INC.
XX
PI    Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI    Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
```

```
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97351.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 175; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 2412 BP; 638 A; 585 C; 551 G; 638 T; 0 other;

     Query Match                 47.3%;   Score 1640.8;   DB 24;   Length 2412;
     Best Local Similarity    99.6%;   Pred. No. 0;
     Matches 1645;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          183 ACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGA 242
                II II IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db          760 ACTAAGACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATGGA 819

Qy          243 AGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCA 302
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db          820 AGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCACCA 879

Qy          303 TCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGAT 362
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db          880 TCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTGAT 939

Qy          363 TTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTT 422
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db          940 TTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTT 999

Qy          423 TCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTA 482
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         1000 TCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTA 1059

Qy          483 AATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTT 542
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         1060 AATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTT 1119

Qy          543 ATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGA 602
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db         1120 ATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGA 1179

Qy          603 ATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCC 662
                IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

```
Db    1180 ATGTCGGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCC 1239

Qy     663 TCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCA 722
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1240 TCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCA 1299

Qy     723 AGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCA 782
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1300 AGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCA 1359

Qy     783 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 842
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1360 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 1419

Qy     843 CTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAA 902
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1420 CTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAA 1479

Qy     903 GCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAA 962
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1480 GCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAA 1539

Qy     963 TCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTT 1022
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1540 TCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTT 1599

Qy    1023 ACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTC 1082
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1600 ACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTC 1659

Qy    1083 CCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGA 1142
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1660 CCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGA 1719

Qy    1143 ATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAAT 1202
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1720 ATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAAT 1779

Qy    1203 TCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCT 1262
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1780 TCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCT 1839

Qy    1263 GTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAA 1322
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1840 GTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAA 1899

Qy    1323 GCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATT 1382
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1900 GCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATT 1959

Qy    1383 GCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACC 1442
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1960 GCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACC 2019

Qy    1443 TCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCT 1502
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2020 TCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCT 2079
```

```
Qy     1503 GTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCC 1562
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2080 GTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCC 2139

Qy     1563 CTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGA 1622
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     2140 CTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTGA 2199

Qy     1623 CTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAGCTCT 1682
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2200 CTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAAAGCTCT 2259

Qy     1683 CAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATC 1742
            ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||||||
Db     2260 CAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAATCAAAATTCCAGATC 2319

Qy     1743 GGTTTCCTGAGAGAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAG 1802
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2320 GGTTTCCTGAGAGAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAG 2379

Qy     1803 GGAGGGCTGGGGGCGGGGAAAAAAAAAAAAAA 1834
            ||||||||||||||||| ||| |||||||||
Db     2380 GGAGGGCTGGGGGCGGGAAAAGAAAAAAAAAA 2411


RESULT 3
AAD46125
ID     AAD46125 standard; DNA; 769 BP.
XX
AC     AAD46125;
XX
DT     27-DEC-2002  (first entry)
XX
DE     Human BAC AC008652 exon used to isolate K+betaM2 cDNA.
XX
KW     Human; potassium channel beta-subunit; K+betaM2 protein; neural disorder;
KW     reproductive disorder; metabolic disorder; premature puberty; nephritis;
KW     endocrine disorder; memory disorder; neuroendocrine condition; asthma;
KW     spermatogenesis; renal disease; learning deficiency; Alzheimer's disease;
KW     neurodegenerative disease; proliferative disorder; autoimmune disease;
KW     carcinoid tumour; blood coagulation disease; blood platelet disease;
KW     rheumatoid arthritis; allergy; hyperproliferative disease; gene therapy;
KW     graft-versus-host disease; organ rejection; antisterility; thrombolytic;
KW     antiinflammatory; neuroprotective; anti-Parkinsonian; immunosuppressive;
KW     nephrotropic; cytostatic; nootropic; hypotensive; vulnerary; ds.
XX
OS     Homo sapiens.
XX
PN     WO200266601-A2.
XX
PD     29-AUG-2002.
XX
PF     24-JAN-2002; 2002WO-US02332.
XX
PR     24-JAN-2001; 2001US-263872P.
PR     14-FEB-2001; 2001US-269794P.
```

XX
PA   (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI   Feder J, Lee L, Chen J, Jackson D, Ramanathan C, Siemers N;
PI   Chang H, Carroll P;
XX
DR   WPI; 2002-691617/74.
XX
PT   New potassium channel beta-subunit, K+betaM2, proteins and nucleic
PT   acids, useful for diagnosing, treating and/or preventing e.g.
PT   reproductive, neural, metabolic, endocrine, memory, neurodegenerative
PT   disorders or diseases -
XX
PS   Example 1; Page 349-350; 366pp; English.
XX
CC   The present invention relates to human potassium channel beta-subunit
CC   (K+betaM2) proteins and polynucleotides encoding such proteins. The
CC   K+betaM2 sequences are useful for diagnosing, treating and/or preventing
CC   reproductive disorders, neural disorders, disorders related to aberrant
CC   potassium regulation or hyper potassium channel activity, metabolic
CC   disorders (e.g. premature puberty), endocrine disorders (e.g. aberrant
CC   growth hormone synthesis and/or secretion), memory disorder, disorders
CC   of the testis (e.g. spermatogenesis), neuroendocrine condition related
CC   to aberrant thyroid hormone release, renal disease or disorders (e.g.
CC   nephritis), disorders related to aberrant higher brain function (e.g.
CC   learning deficiencies), neurodegenerative diseases (e.g. Alzheimer's
CC   disease), proliferative disorders (e.g. carcinoid tumour) and disorders
CC   involving excessive smooth muscle tone or excitability (e.g. asthma).
CC   They may be used to modulate haemostatic or thrombolytic activity, to
CC   treat or prevent blood coagulation diseases or disorders, blood platelet
CC   diseases, wounds, autoimmune diseases, disorders or conditions (e.g.
CC   rheumatoid arthritis), allergic reactions (e.g. asthma), organ rejection
CC   or graft-versus-host disease, and hyperproliferative diseases. K+betaM2
CC   sequences are also used in gene therapy. The present sequence is human
CC   BAC AC008652 exon used to isolate K+betaM2 cDNA. This sequence is used
CC   in the exemplification of the invention.
XX
SQ   Sequence 769 BP; 209 A; 180 C; 184 G; 196 T; 0 other;

     Query Match   22.2%; Score 769; DB 24; Length 769;
     Best Local Similarity   100.0%; Pred. No. 6.5e-143;
     Matches 769; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   393 AGGTCATTTTTTAATAAGTTAGCATCCTTTTCCCTTTCTTACAAGTTGATCCAAAGGATA 452
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGGTCATTTTTTAATAAGTTAGCATCCTTTTCCCTTTCTTACAAGTTGATCCAAAGGATA 60

Qy   453 AGGCTGTGACTCCATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAGAAGAAAGGGA 512
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGGCTGTGACTCCATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAGAAGAAAGGGA 120

Qy   513 CAATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTC 572
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTC 180

Qy   573 CCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCC 632
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db      181 CCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCC 240

Qy      633 ATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAG 692
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 ATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAG 300

Qy      693 ACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGAT 752
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 ACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGAT 360

Qy      753 TCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACT 812
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACT 420

Qy      813 TTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCA 872
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 TTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCA 480

Qy      873 AACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTG 932
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 AACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTG 540

Qy      933 AAGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCG 992
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 AAGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCG 600

Qy      993 ACCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGG 1052
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 ACCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGG 660

Qy     1053 GACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCT 1112
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCT 720

Qy     1113 TGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGA 1161
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGA 769
```

RESULT 4
ABA09216/c

CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.
XX
SQ Sequence 906 BP; 220 A; 225 C; 216 G; 245 T; 0 other;

```
   Query Match            20.2%;  Score 699.2;  DB 22;  Length 906;
   Best Local Similarity  98.9%;  Pred. No. 4.6e-129;
   Matches  704;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;


Qy       515 ATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTCCC 574
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       906 ATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTCCC 847


Qy       575 AACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCCAT 634
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       846 AACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCCAT 787


Qy       635 TCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGAC 694
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       786 TCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGAC 727


Qy       695 ACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGATTC 754
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       726 ACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGATTC 667


Qy       755 TTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTTT 814
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       666 TTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACTTT 607


Qy       815 CCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCAAA 874
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       606 CCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCAAA 547


Qy       875 CTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTGAA 934
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       546 CTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTGAA 487


Qy       935 GATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGAC 994
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       486 GATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGAC 427


Qy       995 CGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGA 1054
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       426 CGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGA 367


Qy      1055 CAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTG 1114
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       366 CAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTG 307
```

```
Qy        1115  GCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAA 1174
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         306  GCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAA 247

Qy        1175  AGATACACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTG 1226
                ||||||||||||||||||||||||||||||||||||||||||    ||   ||
Db         246  AGATACACCTCCAGATTTTATCTCAAATTCAAGCACCTAATGGGGGCACCTG 195
```

RESULT 5
AAS34230
ID   AAS34230 standard; cDNA; 440 BP.
XX
AC   AAS34230;
XX
DT   17-DEC-2001 (first entry)
XX
DE   Human cDNA encoding a novel foetal antigen, SEQ ID No 754.
XX
KW   Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW   immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW   cardiovascular; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; breast neoplasm; cancer;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155312-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01321.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
```

```
PR    14-AUG-2000; 2000US-0225214.
PR    14-AUG-2000; 2000US-0225266.
PR    14-AUG-2000; 2000US-0225267.
PR    14-AUG-2000; 2000US-0225268.
PR    14-AUG-2000; 2000US-0225270.
PR    14-AUG-2000; 2000US-0225447.
PR    14-AUG-2000; 2000US-0225757.
PR    14-AUG-2000; 2000US-0225758.
PR    14-AUG-2000; 2000US-0225759.
PR    18-AUG-2000; 2000US-0226279.
PR    22-AUG-2000; 2000US-0226681.
PR    22-AUG-2000; 2000US-0226868.
PR    22-AUG-2000; 2000US-0227182.
PR    23-AUG-2000; 2000US-0227009.
PR    30-AUG-2000; 2000US-0228924.
PR    01-SEP-2000; 2000US-0229287.
PR    01-SEP-2000; 2000US-0229343.
PR    01-SEP-2000; 2000US-0229344.
PR    01-SEP-2000; 2000US-0229345.
PR    05-SEP-2000; 2000US-0229509.
PR    05-SEP-2000; 2000US-0229513.
PR    06-SEP-2000; 2000US-0230437.
PR    06-SEP-2000; 2000US-0230438.
PR    08-SEP-2000; 2000US-0231242.
PR    08-SEP-2000; 2000US-0231243.
PR    08-SEP-2000; 2000US-0231244.
PR    08-SEP-2000; 2000US-0231413.
PR    08-SEP-2000; 2000US-0231414.
PR    08-SEP-2000; 2000US-0232080.
PR    08-SEP-2000; 2000US-0232081.
PR    12-SEP-2000; 2000US-0231968.
PR    14-SEP-2000; 2000US-0232397.
PR    14-SEP-2000; 2000US-0232398.
PR    14-SEP-2000; 2000US-0232399.
PR    14-SEP-2000; 2000US-0232400.
PR    14-SEP-2000; 2000US-0232401.
PR    14-SEP-2000; 2000US-0233063.
PR    14-SEP-2000; 2000US-0233064.
PR    14-SEP-2000; 2000US-0233065.
PR    21-SEP-2000; 2000US-0234223.
PR    21-SEP-2000; 2000US-0234274.
PR    25-SEP-2000; 2000US-0234997.
PR    25-SEP-2000; 2000US-0234998.
PR    26-SEP-2000; 2000US-0235484.
PR    27-SEP-2000; 2000US-0235834.
PR    27-SEP-2000; 2000US-0235836.
PR    29-SEP-2000; 2000US-0236327.
PR    29-SEP-2000; 2000US-0236367.
PR    29-SEP-2000; 2000US-0236368.
PR    29-SEP-2000; 2000US-0236369.
PR    29-SEP-2000; 2000US-0236370.
PR    02-OCT-2000; 2000US-0236802.
PR    02-OCT-2000; 2000US-0237037.
PR    02-OCT-2000; 2000US-0237038.
PR    02-OCT-2000; 2000US-0237039.
PR    02-OCT-2000; 2000US-0237040.
PR    13-OCT-2000; 2000US-0239935.
```

```
PR    13-OCT-2000; 2000US-0239937.
PR    20-OCT-2000; 2000US-0240960.
PR    20-OCT-2000; 2000US-0241221.
PR    20-OCT-2000; 2000US-0241785.
PR    20-OCT-2000; 2000US-0241786.
PR    20-OCT-2000; 2000US-0241787.
PR    20-OCT-2000; 2000US-0241808.
PR    20-OCT-2000; 2000US-0241809.
PR    20-OCT-2000; 2000US-0241826.
PR    01-NOV-2000; 2000US-0244617.
PR    08-NOV-2000; 2000US-0246474.
PR    08-NOV-2000; 2000US-0246475.
PR    08-NOV-2000; 2000US-0246476.
PR    08-NOV-2000; 2000US-0246477.
PR    08-NOV-2000; 2000US-0246478.
PR    08-NOV-2000; 2000US-0246523.
PR    08-NOV-2000; 2000US-0246524.
PR    08-NOV-2000; 2000US-0246525.
PR    08-NOV-2000; 2000US-0246526.
PR    08-NOV-2000; 2000US-0246527.
PR    08-NOV-2000; 2000US-0246528.
PR    08-NOV-2000; 2000US-0246532.
PR    08-NOV-2000; 2000US-0246609.
PR    08-NOV-2000; 2000US-0246610.
PR    08-NOV-2000; 2000US-0246611.
PR    08-NOV-2000; 2000US-0246613.
PR    17-NOV-2000; 2000US-0249207.
PR    17-NOV-2000; 2000US-0249208.
PR    17-NOV-2000; 2000US-0249209.
PR    17-NOV-2000; 2000US-0249210.
PR    17-NOV-2000; 2000US-0249211.
PR    17-NOV-2000; 2000US-0249212.
PR    17-NOV-2000; 2000US-0249213.
PR    17-NOV-2000; 2000US-0249214.
PR    17-NOV-2000; 2000US-0249215.
PR    17-NOV-2000; 2000US-0249216.
PR    17-NOV-2000; 2000US-0249217.
PR    17-NOV-2000; 2000US-0249218.
PR    17-NOV-2000; 2000US-0249244.
PR    17-NOV-2000; 2000US-0249245.
PR    17-NOV-2000; 2000US-0249264.
PR    17-NOV-2000; 2000US-0249265.
PR    17-NOV-2000; 2000US-0249297.
PR    17-NOV-2000; 2000US-0249299.
PR    17-NOV-2000; 2000US-0249300.
PR    01-DEC-2000; 2000US-0250160.
PR    01-DEC-2000; 2000US-0250391.
PR    05-DEC-2000; 2000US-0251030.
PR    05-DEC-2000; 2000US-0251988.
PR    05-DEC-2000; 2000US-0256719.
PR    06-DEC-2000; 2000US-0251479.
PR    08-DEC-2000; 2000US-0251856.
PR    08-DEC-2000; 2000US-0251868.
PR    08-DEC-2000; 2000US-0251869.
PR    08-DEC-2000; 2000US-0251989.
PR    08-DEC-2000; 2000US-0251990.
PR    11-DEC-2000; 2000US-0254097.
```

CC   The invention relates to novel nucleic acids encoding novel human foetal
CC   antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC   by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC   are also used in diagnosing a pathological condition or susceptibility
CC   to a pathological condition. The antibodies to the antigens can also
CC   be used in alleviating symptoms associated with the disorders and in
CC   diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC   include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC   and ocular disorders e.g. corneal infection. The polypeptides can also
CC   be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. Numerous
CC   examples of diseases and disorders treated by the nucleic acids and
CC   proteins are given in the specification. The present sequence

   Query Match   12.2%; Score 423.4; DB 22; Length 440;
   Best Local Similarity 98.9%; Pred. No. 1.7e-74;
   Matches 435; Conservative   0; Mismatches   4; Indels   1; Gaps   1;

Qy   1842 TTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGAT 1901
            ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   1 TTTGAAATTAACCTCCTAAAAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGAT 60

Qy   1902 GCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCAC 1961
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCAC 120

Qy   1962 CTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTT 2021
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTT 180

```
Qy   2022 TTTTAGTTATTTGTTTGTTTACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAG 2081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTTAGTTATTTGTTTGTTTACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAG 240

Qy   2082 CCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGG-TTTTT 2140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    241 CCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTT 300

Qy   2141 TCTCTCATCCTTCTACCTCCCTCCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCA 2200
          ||||||||||||||||||    |||||||||||||||||||||||||||||||||||||||
Db    301 TCTCTCATCCTTCTACCTNNCTCCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCA 360

Qy   2201 ATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGA 2260
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    361 ATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTTCCTTTTGTNTATGGGGTTGGGGGGA 420

Qy   2261 ATGGCAGATTTATATGACTT 2280
          ||||||||||||||||||||
Db    421 ATGGCAGATTTATATGACTT 440
```

RESULT 6
AAD49513
ID   AAD49513 standard; cDNA; 2398 BP.
XX
AC   AAD49513;
XX
DT   24-MAR-2003 (first entry)
XX
DE   Human TRICH-15 cDNA.
XX
KW   Human; transporter and ion channel; TRICH; atherosclerosis; cancer;
KW   gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   114..1535
FT   /*tag= a
FT   /product= "Human TRICH protein"
FT   sig_peptide   114..230
FT   /*tag= b
FT   mat_peptide   231..1532
FT   /*tag= c
FT   /product= "Mature human TRICH protein"
XX
PN   WO200283712-A2.
XX
PD   24-OCT-2002.
XX
PF   12-APR-2002; 2002WO-US11760.
XX
PR   12-APR-2001; 2001US-283440P.
PR   20-APR-2001; 2001US-285592P.
PR   27-APR-2001; 2001US-287263P.
PR   04-MAY-2001; 2001US-288666P.
PR   18-MAY-2001; 2001US-292042P.

```
   Query Match                9.2%;  Score 319.4;  DB 25;   Length 2398;
   Best Local Similarity     58.0%;  Pred. No. 9.9e-54;
   Matches  769;  Conservative    0;  Mismatches  466;  Indels   90;  Gaps    8;

Qy       560 GGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTT 619
             ||| ||    |   |||    ||||||||||| || || ||||||||| || || || || |||
Db       216 GGGCCCTGCGCACCCTCGCCCTTCCCTGAAGTAGTGGAGCTGAACGTAGGCGGCCAGGTT 275

Qy       620 TATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTT 679
             ||| | ||    || || ||  || | ||| ||||  |        ||    |   ||||| 
Db       276 TATGTGACCAAGCACTCGACGCTGCTCAGCGTCCCGGACAGTACTTTGGCCAGCATGTTC 335

Qy       680 TCCCCAAAGAGAGACACGGCTAAT--------------GATCTAGCCAAGGACTCCAAG 724
             || ||     || |   |            || || ||| |||| | |
Db       336 TCGCCCTCTAGTCCCCGTGGCGGCGCCCGGCGCCGGGGCGAGCTGCCCAGGGACAGCCGG 395

Qy       725 GGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGG 784
             |     | || ||||| ||| | || || ||| | ||| | || | || | ||||| ||||| ||
Db       396 GCGCGCTTCTTCATCGACCGGGACGGCTTCCTTTTCAGGTACGTGCTGGATTATCTGCGG 455

Qy       785 GACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCT 844
             |||| |||  | |  ||||| || ||||| || || || | | |||    | || ||
Db       456 GACAAGCAACTCGCGCTGCCGGAGCACTTCCCCGAGAAGGAGCGGCTGCTGCGCGAGGCC 515

Qy       845 GAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGC 904
             || || |||||||||| | ||||||||||| || ||| | ||| | | | ||||||| | |
```

```
Db    516 GAGTATTTCCAGCTCACCGACTTGGTCAAGCTGCTGTCGCCCAAGGTCACCAAGCAGAAC 575

Qy    905 CC------AGATGAATTCTGCCACAGTGACTTTGA------------------AGATGC 939
              |     || ||   |||||| || ||| | ||                  || ||
Db    576 TCTCTCAACGACGAGGGCTGCCAGAGCGACCTGGAGGACAACGTCTCGCAGGGTAGCAGC 635

Qy    940 CTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCC------------ 986
              || |   |  || | |  |       | || || | | | |  ||
Db    636 GACGCGCTGCTGCTGCGCGGGGCGGCGGCCGCCGTGCCCTCGGGCCCGGGAGCGCACGGT 695

Qy    987 ------------------CTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAGA 1027
                           |    |||      | ||| ||| | ||  |||| ||| |
Db    696 GGTGGCGGCGGCGGCGCGCAGGACAAGCGCTCGGGCTTCCTCACGCTGGGCTACCGG 755

Qy   1028 GGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGG 1087
              || |||| ||||    |   | ||    ||||| || ||||| || ||| | ||  | ||
Db    756 GGCTCCTACACCACCGTGCGCGACAACCAGGCCGACGCCAAATTCCGGCGTGTGGCGCGC 815

Qy   1088 ATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAA 1147
              ||   |||| || ||   | ||  |    |||| || || ||||| || || ||  | || ||
Db    816 ATCATGGTGTGCGGGCGCATCGCGCTGGCCAAGGAGGTCTTCGGGGACACGCTCAACGAG 875

Qy   1148 AGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAAG 1207
              ||| | ||||| || ||   || || |  ||||| ||| | || || ||||| ||||
Db    876 AGCCGCGACCCCGACCGGCAGCCGGAGAAGTACACGTCCCGCTTCTACCTCAAGTTCACC 935

Qy   1208 CACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAAC 1267
              || ||||    ||| ||||||      |||| |||   || |||||||||||||||| ||||||
Db    936 TACTTGGAGCAGGCCTTTGATCGCCTGTCCGAGGCCGGCTTCCACATGGTGGCGTGTAAC 995

Qy   1268 TCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTCAAGCTAC 1327
              || ||||  || ||   | ||| ||||||| ||    || ||||||||||||||||   ||||||
Db    996 TCCTCGGGCACCGCCGCCTTCGTCAACCAGTACCGCGACGACAAGATCTGGAGCAGCTAC 1055

Qy   1328 ACTGAATATGTCTTCTACCGTGAGCCT---TCCAGATGGTCACCCTCACACTGCGATTGC 1384
              || || ||  | |||| |||   |||        || | | ||||| |||    ||
Db   1056 ACCGAGTACATTTTCTTCCGACCACCTCAGAAAATAGTATCACCTAAACAAGAACATGAA 1115

Qy   1385 TGCTGCAAGAATGGCAAAG---GTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGAC 1441
               | ||   ||| |||||     ||| ||||| | | || || || || || || || |||||
Db   1116 GATAGGAAACATGACAAAGTCACTGATAAAGGAAGTGAAAGTGGGACTTCCTGTAATGAG 1175

Qy   1442 CTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATC 1501
              |||||||| || || || | |||||||| || ||||| ||| ||||||||||||| |      |
Db   1176 CTCTCCACTTCCAGTTGTGACAGCCATTCAGAGGCAAGCACTCCCCAGGACAACCCATCC 1235

Qy   1502 TGTGGTCCCGTGACA------CGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAG 1555
              ||| |     | ||     | ||| | |||    || |||| || ||| ||
Db   1236 AGTGCCCAGCAGGCAACAGCTCACCAACCTAACACTTTAACATTGGATCGCCCCTCTAAA 1295

Qy   1556 AAGGGCCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGG 1615
              ||| | ||||| ||   |||   |  |||| | || | || || || || | | ||| |
Db   1296 AAAGCACCTGTACAATGGATACCCCCACCAGACAAACGCAGAAACAGTGAACTCTTTCAG 1355

Qy   1616 ATTCTGACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCTGTTAAAGAA 1675
              |  || |      ||| ||||| | || || || || || || |||         ||    ||
Db   1356 ACCCTCATCAGCAAGTCCCGGGAAACAAATCTGTCCAAAAAGAAA------GTCTGTGAG 1409
```

```
Qy   1676 AAGCTCTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAATT 1735
          |||||       | || || ||   || | || ||||| |||||||||  |||||   | |||
Db   1410 AAGCTAAGTGTGGAAGAAGAAATGAAAAGTGTATTCAGGATTTTAAAAAAATCCACATT 1469

Qy   1736 CCAGATCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCAT 1795
          ||||||      ||||| ||| | ||||  |  |||||||||||||| ||    |||||||
Db   1470 CCAGATTATTTTCCAGAGCGCAAACGCCAATGGCAATCTGAACTGTTGCAGAAGTATGGG 1529

Qy   1796 CTATA 1800
          ||||
Db   1530 TTATA 1534


RESULT 7
ABV99059
ID   ABV99059 standard; cDNA; 632 BP.
XX
AC   ABV99059;
XX
DT   14-JAN-2003   (first entry)
XX
DE   Human pancreatic cancer expressed cDNA SEQ ID NO 4467.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200260317-A2.
XX
PD   08-AUG-2002.
XX
PF   30-JAN-2002; 2002WO-US02781.
XX
PR   30-JAN-2001; 2001US-265305P.
PR   31-JAN-2001; 2001US-265682P.
PR   09-FEB-2001; 2001US-267568P.
PR   21-MAR-2001; 2001US-278651P.
PR   28-APR-2001; 2001US-287112P.
PR   16-MAY-2001; 2001US-291631P.
PR   12-JUL-2001; 2001US-305484P.
PR   20-AUG-2001; 2001US-313999P.
PR   27-NOV-2001; 2001US-333626P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Benson DR,  Kalos MD,  Lodes MJ,  Persing DH,  Hepler WT,  Jiang Y;
XX
DR   WPI; 2002-627435/67.
XX
PT   New isolated polynucleotide and pancreatic tumor polypeptides, useful
PT   for diagnosing, preventing and/or treating cancer, particularly
PT   pancreatic cancer  -
XX
PS   Claim 1; SEQ ID NO 4467; 300pp + Sequence Listing; English.
XX
```

CC   The invention relates to an isolated polynucleotide (I) comprising: (a)
CC   any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145);
CC   (b) complements of (a); (c) sequences consisting of at least 20
CC   contiguous residues of (a); (d) sequences that hybridize to (a), under
CC   moderately stringent conditions; (e) sequences having at least 75% or 90%
CC   identity to (a); or (f) degenerate variants of (a). Polypeptides
CC   (ABP68596-ABP68637) encoded by (I) and oligonucleotide can be used to
CC   detect cancer in a patient and compositions comprising polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations and
CC   antigen presenting cells expressing the polypeptide are useful in
CC   treating pancreatic cancer and stimulating an immune response. The
CC   polynucleotides can be used as probes or primers for nucleic acid
CC   hybridisation, in the design and preparation of ribozyme molecules for
CC   inhibiting expression of the tumour polypeptides and proteins in the
CC   tumour cells, in vaccines and for gene therapy.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 632 BP; 178 A; 92 C; 119 G; 239 T; 4 other;

     Query Match   5.9%;   Score 205;   DB 24;   Length 632;
     Best Local Similarity   93.4%;   Pred. No. 3.4e-31;
     Matches   214; Conservative   0; Mismatches   15; Indels   0; Gaps   0;

Qy   3206 AGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATT 3265
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15 AGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATT 74

Qy   3266 TTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGT 3325
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 TTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGT 134

Qy   3326 GTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACA 3385
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135 GTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACA 194

Qy   3386 GGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAA 3434
               ||||||||||||||||||||||||||| | | |||| || | ||
Db   195 GGTAATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA 243

RESULT 8
ABV95156
ID   ABV95156 standard; cDNA; 614 BP.
XX
AC   ABV95156;
XX
DT   14-JAN-2003 (first entry)
XX
DE   Human pancreatic cancer expressed cDNA SEQ ID NO 564.
XX
KW   Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW   cytostatic; tumour; gene; ss.
XX
OS   Homo sapiens.
XX

```
PN    WO200260317-A2.
XX
PD    08-AUG-2002.
XX
PF    30-JAN-2002; 2002WO-US02781.
XX
PR    30-JAN-2001; 2001US-265305P.
PR    31-JAN-2001; 2001US-265682P.
PR    09-FEB-2001; 2001US-267568P.
PR    21-MAR-2001; 2001US-278651P.
PR    28-APR-2001; 2001US-287112P.
PR    16-MAY-2001; 2001US-291631P.
PR    12-JUL-2001; 2001US-305484P.
PR    20-AUG-2001; 2001US-313999P.
PR    27-NOV-2001; 2001US-333626P.
XX
PA    (CORI-) CORIXA CORP.
XX
PI    Benson DR,  Kalos MD,  Lodes MJ,  Persing DH,  Hepler WT,  Jiang Y;
XX
DR    WPI; 2002-627435/67.
XX
PT    New isolated polynucleotide and pancreatic tumor polypeptides, useful
PT    for diagnosing, preventing and/or treating cancer, particularly
PT    pancreatic cancer  -
XX
PS    Claim 1; SEQ ID NO 564; 300pp + Sequence Listing; English.
XX
CC    The invention relates to an isolated polynucleotide (I) comprising: (a)
CC    any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145);
CC    (b) complements of (a); (c) sequences consisting of at least 20
CC    contiguous residues of (a); (d) sequences that hybridize to (a), under
CC    moderately stringent conditions; (e) sequences having at least 75% or 90%
CC    identity to (a); or (f) degenerate variants of (a). Polypeptides
CC    (ABP68596-ABP68637) encoded by (I) and oligonucleotide can be used to
CC    detect cancer in a patient and compositions comprising polypeptides,
CC    polynucleotides, antibodies, fusion proteins, T cell populations and
CC    antigen presenting cells expressing the polypeptide are useful in
CC    treating pancreatic cancer and stimulating an immune response. The
CC    polynucleotides can be used as probes or primers for nucleic acid
CC    hybridisation, in the design and preparation of ribozyme molecules for
CC    inhibiting expression of the tumour polypeptides and proteins in the
CC    tumour cells, in vaccines and for gene therapy.
CC    Note: The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 614 BP; 177 A; 87 C; 110 G; 236 T; 4 other;

      Query Match              5.8%;  Score 201;  DB 24;  Length 614;
      Best Local Similarity   93.3%;  Pred. No. 2.1e-30;
      Matches  210; Conservative    0; Mismatches   15; Indels   0; Gaps    0;

Qy       3210 ACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATTTTCA 3269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATTTTCA 60
```

```
Qy    3270 GCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGC 3329
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGC 120

Qy    3330 CAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTA 3389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTA 180

Qy    3390 ATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAA 3434
           ||||||||||||||||||||     |  |  |||| ||  |  ||
Db     181 ATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA 225
```

RESULT 9
ABT09812

```
ID    ABT09812 standard; cDNA; 2052 BP.
XX
AC    ABT09812;
XX
DT    05-DEC-2002  (first entry)
XX
DE    Polynucleotide encoding the K+beta M6 protein SEQ ID No 1.
XX
KW    Cytostatic; cardiant; neuroprotective; immunomodulator; antimigraine;
KW    sedative; gynaecological;; potassium channel beta subunit; K+betaM6;
KW    gastrointestinal; reproductive; neural; sleep; low DNA repair capacity;
KW    hyperpotassium channel activity; cardiovascular; melatonin synthesis;
KW    mammary cancer tumourigenesis; pineal gland associated disorder;
KW    pulmonary disorder; immune disorder; NF-kB activity; migraine headache;
KW    low free-radical buffering capacity; delayed sleep phase syndrome;
KW    circadian cycle; melatonin secretion; cancer; gene; ss.
XX
OS    Homo sapiens.
XX
PN    WO200270727-A2.
XX
PD    12-SEP-2002.
XX
PF    21-FEB-2002; 2002WO-US05674.
XX
PR    21-FEB-2001; 2001US-270132P.
PR    27-MAR-2001; 2001US-278953P.
XX
PA    (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI    Feder J,   Lee L,   Chen J,   Jackson DG,   Ramanathan C,   Siemers N;
PI    Chang H;
XX
DR    WPI; 2002-713455/77.
DR    P-PSDB; ABJ10886.
XX
PT    New polynucleotide encoding human potassium channel beta subunit
PT    polypeptide, useful for diagnosing, preventing, treating or
PT    ameliorating e.g. cancer  -
XX
PS    Claim 1; Fig 1; 332pp; English.
XX
```

CC   The invention relates to an isolated polynucleotide encoding a potassium
CC   channel beta subunit (K+betaM6) polypeptide or its variants. The human
CC   potassium beta subunit polynucleotide or polypeptide is useful for
CC   diagnosing, preventing, treating or ameliorating a pathological condition
CC   such as gastrointestinal, reproductive, neural, sleep, cardiovascular or
CC   pulmonary disorders, a disorder related to hyperpotassium channel
CC   activity, an immune disorder related to aberrant NF-kB activity, pineal
CC   gland associated disorders, migraine headaches, disorders associated with
CC   aberrant melatonin synthesis and/or release or with low DNA repair
CC   capacities or low free-radical buffering capacity, delayed sleep phase
CC   syndrome, aberrations in circadian cycle, mammary cancer tumourigenesis,
CC   age related disorders associated with decreased melatonin secretion, or
CC   cancer. This polynucleotide sequence represents the cDNA encoding the
CC   potassium channel beta subunit (K+betaM6) protein of the invention.
XX
SQ   Sequence 2052 BP; 380 A; 640 C; 607 G; 425 T; 0 other;

   Query Match   4.8%;   Score 167;   DB 24;   Length 2052;
   Best Local Similarity   64.6%;   Pred. No. 1.5e-23;
   Matches 267;   Conservative   0;   Mismatches 140;   Indels   6;   Gaps   1;

Qy   967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
                 | | || ||| ||| | | ||| | ||| | ||| || | || ||| |
Db   705 CACGCCGTCCCAGTCGCTGGACGGCAGCCGGCGCTCGGGCTACATCACCATCGGCTACCG 764

Qy   1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
                 || ||||| ||||| | || | ||| | |||||| || |||||||| ||| |||| | ||
Db   765 CGGCTCCTACACCATCGGGCGGGACGCGCAGGCGGACGCCAAGTTCCGGCGAGTGGCGCG 824

Qy   1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
                 || ||||| |||| || || |||| || || || ||||| || || |||| ||
Db   825 CATCACCGTTTGCGGAAAGACGTCGCTGGCCAAGGAGGTGTTTGGGGACACCCTGAACGA 884

Qy   1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
                 |||| | ||||| || || |||| || | |||||||| | | ||| ||||| |||||
Db   885 AAGCCGGGACCCCGACCGTCCCCCGGAGCGCTACACCTCGCGCTATTACCTCAAGTTCAA 944

Qy   1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
                 ||||| | ||| || || || | | |||| |||| || |||||||||||||| || |
Db   945 CTTCCTGGAGCAGGCCTTCGACAAGCTGTCCGAGTCGGGCTTCCACATGGTGGCGTGCAG 1004

Qy   1267 CTCATCGGTGACAGCATCTTT------CATCAACCAATATACAGATGACAAGATCTGGTC 1320
                 ||| ||| || | || || || | | | || || ||||||||||||| |
Db   1005 CTCCACGGGCACCTGCGCCTTTGCCAGCAGCACCGACCAGAGCGAGGACAAGATCTGGAC 1064

Qy   1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCAC 1373
                 ||||||||| || || ||||||| | | ||| | | |||||| |
Db   1065 CAGCTACACCGAGTACGTCTTCTGCAGGGAGTGAGCTCCCCAGACCCCTCGC 1117

RESULT 10
ABQ88125/c
ID   ABQ88125 standard; cDNA; 109201 BP.
XX
AC   ABQ88125;
XX
DT   18-SEP-2002   (first entry)

```
XX
DE    Human osteoblast differentiation related cDNA SEQ ID NO 32.
XX
KW    Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW    osteoporosis; osteopathic; ss.
XX
OS    Homo sapiens.
XX
PN    WO200250301-A2.
XX
PD    27-JUN-2002.
XX
PF    18-DEC-2001; 2001WO-US48276.
XX
PR    18-DEC-2000; 2000US-255882P.
PR    24-APR-2001; 2001US-285691P.
XX
PA    (GENE-) GENE LOGIC INC.
PA    (PROC ) PROCTER & GAMBLE CO.
XX
PI    Ji D,  Axelrod DW,  Cook JS,  Jaiswal N,  Einstein R,  Houghton A;
PI    Mertz L;
XX
DR    WPI; 2002-557663/59.
XX
PT    Use of genes and their expression profiles associated with osteoblast
PT    differentiation for screening modulators bone formation, for diagnosing
PT    or treating e.g. osteoporosis, or as markers for the differentiation
PT    process  -
XX
PS    Claim 1; SEQ ID NO 32; 78pp + Sequence Listing; English.
XX
CC    The invention relates to genes and their expression profiles are used
CC    for:
CC    (a) screening modulators of precursor stem cell differentiation into
CC    osteoblasts, or bone tissue deposition;
CC    (b) diagnosing abnormal deposition of bone tissue, abnormal rate of
CC    osteoblast formation or osteoporosis; or
CC    (c) treating or monitoring treatment of the conditions cited in (b), or
CC    monitoring the progression of bone tissue deposition.
CC    Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC    osteoporosis or male osteoporosis, osteopenia, osteodystrophy,
CC    drug-induced abnormalities in bone formation or bone loss, conditions
CC    that involve altered bone metabolism (e.g. idiopathic juvenile
CC    osteoporosis), skeletal disease linked to breast cancer, mastocytosis,
CC    Fanconi syndrome or fibrous dysplasia. The present sequence is that of an
CC    osteoblast differentiation associated cDNA marker of the invention.
CC    Note: The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ    Sequence 109201 BP; 32871 A; 23488 C; 22108 G; 30734 T; 0 other;

   Query Match             4.8%;  Score 167;  DB 24;  Length 109201;
   Best Local Similarity   64.6%;  Pred. No. 3.5e-23;
   Matches  267;  Conservative    0;  Mismatches  140;  Indels    6;  Gaps    1;
```

```
Qy      967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
            | |  || |||      |||     | |  |||   | ||| | ||| ||  | || |||| |
Db     9291 CACGCCGTCCCAGTCGCTGGACGGCAGCCGGCGCTCGGGCTACATCACCATCGGCTACCG 9232

Qy     1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
            || |||| |||| | ||  | ||| | ||||| || |||||||| ||| |||| | || 
Db     9231 CGGCTCCTACACCATCGGGCGGGACGCGCAGGCGGACGCCAAGTTCCGGCGAGTGGCGCG 9172

Qy     1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
            ||     ||||| |||| ||  ||  |||| || || || ||||| || ||  |||| ||
Db     9171 CATCACCGTTTGCGGAAAGACGTCGCTGGCCAAGGAGGTGTTTGGGGACACCCTGAACGA 9112

Qy     1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
            |||| | ||||| || || |||| ||  | ||||||||| | | ||| ||||| |||||
Db     9111 AAGCCGGGACCCCGACCGTCCCCCGGAGCGCTACACCTCGCGCTATTACCTCAAGTTCAA 9052

Qy     1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
            |||||| ||| || || || || | | |||| ||||  || |||||||||||||| || |
Db     9051 CTTCCTGGAGCAGGCCTTCGACAAGCTGTCCGAGTCGGGCTTCCACATGGTGGCGTGCAG 8992

Qy     1267 CTCATCGGTGACAGCATCTTT------CATCAACCAATATACAGATGACAAGATCTGGTC 1320
            ||| ||| || | || || || || | | | || |||||||||||| |
Db     8991 CTCCACGGGCACCTGCGCCTTTGCCAGCAGCACCGACCAGAGCGAGGACAAGATCTGGAC 8932

Qy     1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCAC 1373
            |||||||| || ||  ||||||| | |  |||      | | ||||| |
Db     8931 CAGCTACACCGAGTACGTCTTCTGCAGGGAGTGAGCTCCCCAGACCCCCTCGC 8879


RESULT 11
ABQ40654
ID    ABQ40654 standard; DNA; 854 BP.
XX
AC    ABQ40654;
XX
DT    12-JUL-2002   (first entry)
XX
DE    Oligonucleotide for detecting cytosine methylation SEQ ID NO 27245.
XX
KW    Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW    drug; side effect; cancer; central nervous system; cardiovascular;
KW    gastrointestinal; respiratory system; single nucleotide polymorphism;
KW    SNP; cell differentiation; ds.
XX
OS    Homo sapiens.
XX
PN    WO200218632-A2.
XX
PD    07-MAR-2002.
XX
PF    01-SEP-2001; 2001WO-EP10074.
XX
PR    01-SEP-2000; 2000DE-1043826.
PR    05-SEP-2000; 2000DE-1044543.
XX
PA    (EPIG-) EPIGENOMICS AG.
XX
```

PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 854 BP; 131 A; 98 C; 289 G; 336 T; 0 other;

      Query Match   3.3%;   Score 114.2; DB 24;   Length 854;
      Best Local Similarity   59.7%;   Pred. No. 3.6e-13;
      Matches 213; Conservative   0; Mismatches 138; Indels   6; Gaps   1;

Qy   1001 TGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCA 1060
                 | ||||| |||| | | |||||| | || | | | | || | || | ||||
Db   491 TCGGGTTATATTATTATCGGTTATCGCGGTTTTTATATTATCGGGCGGGACGCGTAGGCG 550

Qy   1061 GATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAA 1120
                || | ||||||||| |||| | || ||| | ||||| |||| || || |||| ||
Db   551 GACGTTAAGTTTCGGCGAGTGGCGCGTATTATCGTTTGCGGAAAGACGTCGTTGGTTAAG 610

Qy   1121 GAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATAC 1180
                || || ||||| || | |||||| ||||| | || |||||| | || | ||
Db   611 GAGGTGTTTGGGGATATTTTGAACGAAAGTCGGGATTTCGATCGTTTTTCGGAGCGTTAT 670

Qy   1181 ACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAG 1240
                 | || | | |||| | || || || |||| || ||| |||| ||||| |||
Db   671 ATTTCGCGTTATTATTTTAAGTTTAATTTTTTGGAGTAGGTTTTCGATAAGTTGTTCGAG 730

Qy   1241 TGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTT------CATCAAC 1294
                | || || | ||||||||| |||| | ||| | || | | |
Db   731 TCGGGTTTTTATATGGTGGCGTGTAGTTTTACGGGTATTTGCGTTTTTGTTAGTAGTATC 790

```
Qy      1295 CAATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAG 1351
              |  ||  |   ||  ||  |||||  |||    ||  ||  |  ||  ||  ||  ||  |    |  |||
Db       791 GATTAGAGCGAGGATAAGATTTGGATTAGTTATATCGAGTACGTTTTTTGTAGGGAG 847
```

RESULT 12
ABQ40655/c
ID   ABQ40655 standard; DNA; 854 BP.
XX
AC   ABQ40655;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 27246.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide

```
  Query Match                 3.3%;   Score 114.2;   DB 24;   Length 854;
  Best Local Similarity   59.7%;   Pred. No. 3.6e-13;
  Matches  213;  Conservative    0;  Mismatches  138;  Indels    6;  Gaps    1;

Qy      1001 TGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCA 1060
             | ||||| |||| | | ||||| | || | |  |   | || | || |   ||||
Db       364 TCGGGTTATATTATTATCGGTTATCGCGGTTTTTATATTATCGGGCGGGACGCGTAGGCG 305

Qy      1061 GATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAA 1120
             || | ||||||||| |||| | || ||| | ||||| |||| || || |||| ||
Db       304 GACGTTAAGTTTCGGCGAGTGGCGCGTATTATCGTTTGCGGAAAGACGTCGTTGGTTAAG 245

Qy      1121 GAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATAC 1180
             || || |||||| || | ||||||| ||||| | ||     |||||   | || | ||
Db       244 GAGGTGTTTGGGGATATTTTGAACGAAAGTCGGGATTTCGATCGTTTTTCGGAGCGTTAT 185

Qy      1181 ACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAG 1240
             |  ||  | | ||||| | || || ||    |||| || ||| |||| ||||| |||
Db       184 ATTTCGCGTTATTATTTTAAGTTTAATTTTTTGGAGTAGGTTTTCGATAAGTTGTTCGAG 125

Qy      1241 TGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTT------CATCAAC 1294
             |  || || | ||||||||| |||| | ||| |   |||   | | |
Db       124 TCGGGTTTTTATATGGTGGCGTGTAGTTTTACGGGTATTTGCGTTTTTGTTAGTAGTATC 65

Qy      1295 CAATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAG 1351
             | || |  || || ||||| |||  || || | | || || || || || |  | |||
Db        64 GATTAGAGCGAGGATAAGATTTGGATTAGTTATATCGAGTACGTTTTTTGTAGGGAG 8
```

PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 1757 BP; 246 A; 209 C; 640 G; 662 T; 0 other;

      Query Match   3.2%; Score 109.8; DB 24; Length 1757;
      Best Local Similarity   62.6%; Pred. No. 3.1e-12;
      Matches   171; Conservative   0; Mismatches 102; Indels   0; Gaps   0;

Qy   1001 TGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCA 1060
             | ||||| |||| | | ||||| | || | | | | || | || | ||||
Db   1485 TCGGGTTATATTATTATCGGTTATCGCGGTTTTTATATTATCGGGCGGGACGCGTAGGCG 1544

Qy   1061 GATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAA 1120
             || | ||||||||| |||| | || ||| | ||||| |||| || || |||| ||
Db   1545 GACGTTAAGTTTCGGCGAGTGGCGCGTATTATCGTTTGCGGAAAGACGTCGTTGGTTAAG 1604

Qy   1121 GAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATAC 1180
             || || |||||| || | ||||||| ||||| | || ||||| | || | ||
Db   1605 GAGGTGTTTGGGGATATTTTGAACGAAAGTCGGGATTTCGATCGTTTTTCGGAGCGTTAT 1664

```
Qy      1181 ACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAG 1240
             |  || | | |||| | || || ||     |||| || ||| |||| ||||| |||
Db      1665 ATTTCGCGTTATTATTTTAAGTTTAATTTTTTGGAGTAGGTTTTCGATAAGTTGTTCGAG 1724

Qy      1241 TGTGGATTCCACATGGTGGCCTGTAACTCATCG 1273
             |  || || | |||||||| |||| |   ||
Db      1725 TCGGGTTTTTATATGGTGGCGTGTAGTTTTACG 1757
```

RESULT 14
ABQ13669/c
ID   ABQ13669 standard; DNA; 1757 BP.
XX
AC   ABQ13669;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 260.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
```

```
  Query Match              3.2%;   Score 109.8;   DB 24;   Length 1757;
  Best Local Similarity   62.6%;   Pred. No. 3.1e-12;
  Matches  171;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;


Qy        1001 TGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCA 1060
               I IIIII  IIII I I  IIIII  I II I  I  I   I II  I II I   IIII
Db         273 TCGGGTTATATTATTATCGGTTATCGCGGTTTTTATATTATCGGGCGGGACGCGTAGGCG 214


Qy        1061 GATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAA 1120
               II I  IIIIIIIII IIII  I II III I IIIII IIII II  II IIII  II
Db         213 GACGTTAAGTTTCGGCGAGTGGCGCGTATTATCGTTTGCGGAAAGACGTCGTTGGTTAAG 154


Qy        1121 GAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATAC 1180
               II II IIIII II I IIIIII IIIII  I II     IIIII    I II  I II
Db         153 GAGGTGTTTGGGGATATTTTGAACGAAAGTCGGGATTTCGATCGTTTTTCGGAGCGTTAT 94


Qy        1181 ACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAG 1240
               I  II I I IIII I II II II      IIII   II III IIII IIIII  III
Db          93 ATTTCGCGTTATTATTTTAAGTTTAATTTTTTGGAGTAGGTTTTCGATAAGTTGTTCGAG 34


Qy        1241 TGTGGATTCCACATGGTGGCCTGTAACTCATCG 1273
               I  II II  I IIIIIIIII IIII  I   II
Db          33 TCGGGTTTTTATATGGTGGCGTGTAGTTTTACG 1
```

RESULT 15
ABQ40656/c

XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 854 BP; 132 A; 98 C; 302 G; 322 T; 0 other;

     Query Match   3.1%; Score 108.6; DB 24; Length 854;
     Best Local Similarity   57.1%; Pred. No. 4.6e-12;
     Matches 220; Conservative   0; Mismatches 159; Indels   6; Gaps   1;

Qy   967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
              I I II III   III   I I IIII   I I   I III II I I III I
Db   398 CACGCCGTCCCAATCGCTAAACGACAACCGACGCTCGAACTACATCACCATCGACTACCG 339

Qy   1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
              I IIII IIII I I   II I I II I I IIIII II II II I I II
Db   338 CGACTCCTACACCATCGAACGAAACGCGCAAACGAACGCCAAATTCCGACGAATAACGCG 279

Qy   1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
              II   IIII I II I II I I III II I III I I II I II II

```
Db       278 CATCACCGTTTACGAAAAAACGTCGCTAACCAAAAAAATATTTAAAAACACCCTAAACGA 219

Qy      1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
             || | || |||| || ||  ||||  || | ||||||||  | | ||| |||||||||||
Db       218 AAACCGAAACCCCGACCGTCCCCCGAAACGCTACACCTCGCGCTATTACCTCAAATTCAA 159

Qy      1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
             |||  ||    | || || |   | || || |      |||||||||  |   | |  ||
Db       158 CTTCCTAAAACAAACCTTCGACAAACTATCCGAATCGAACTTCCACATAATAACGTACAA 99

Qy      1267 CTCATCGGTGACAGCATCTTT------CATCAACCAATATACAGATGACAAGATCTGGTC 1320
             |||  ||   ||     | ||      || || |  | |   ||  |||| |||| |||  |
Db        98 CTCCACGAACACCTACGCCTTTACCAACAACACCGACCAAAACGAAAACAAATCTAAAC 39

Qy      1321 AAGCTACACTGAATATGTCTTCTAC 1345
              | |||||| ||||| |||||||||
Db        38 CAACTACACCGAATACGTCTTCTAC 14


Search completed: January 28, 2004, 21:07:41
Job time : 872 secs
```

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 20:50:10 ; Search time 192 Seconds
                                             (without alignments)
                                             7972.483 Million cell updates/sec

Title: US-10-056-884A-1
Perfect score: 3468
Sequence: 1 caagcactgtgctaaagtgt..........aaaaaaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                    SUMMARIES
                            %
Result   Query
   No.   Score Match Length DB   ID   Description
----
      1   74   2.1   664   4   US-09-904-615-66   Sequence 66, Appl
      2   73.8   2.1   1091   4   US-09-328-965-1   Sequence 1, Appli
c   3   73.4   2.1   4055   4   US-09-620-312D-706   Sequence 706, App
      4   71.2   2.1   1701   4   US-09-996-243-114   Sequence 114, App
      5   69.2   2.0   1441   3   US-08-821-994-63   Sequence 63, Appl
      6   69.2   2.0   2246   4   US-09-363-708-3   Sequence 3, Appli
      7   69.2   2.0   2246   4   US-09-083-587-3   Sequence 3, Appli
      8   69.2   2.0   2406   4   US-09-594-506-37   Sequence 37, Appl
      9   69   2.0   2202   4   US-09-465-558-59   Sequence 59, Appl
     10   68.8   2.0   1147   1   US-08-665-716-1   Sequence 1, Appli
     11   68.6   2.0   1736   3   US-09-182-816-22   Sequence 22, Appl

```
c  12    68.6    2.0    1736   3   US-09-182-816-24        Sequence 24, Appl
   13    68.6    2.0    1736   3   US-09-471-528-22        Sequence 22, Appl
c  14    68.6    2.0    1736   3   US-09-471-528-24        Sequence 24, Appl
   15    68.6    2.0    1736   3   US-09-634-530-22        Sequence 22, Appl
c  16    68.6    2.0    1736   3   US-09-634-530-24        Sequence 24, Appl
   17    68.4    2.0     593   4   US-09-904-615-59        Sequence 59, Appl
   18    68.4    2.0    2323   4   US-09-149-476-24        Sequence 24, Appl
   19    68.4    2.0    2806   4   US-09-653-839-9         Sequence 9, Appli
   20    68.4    2.0    3848   3   US-09-112-096-28        Sequence 28, Appl
   21    68.4    2.0    5668   3   US-09-112-096-14        Sequence 14, Appl
   22    68      2.0    1098   3   US-09-248-335-35        Sequence 35, Appl
   23    68      2.0    2447   2   US-09-014-969-14        Sequence 14, Appl
   24    67.6    1.9    5503   2   US-08-726-012B-1        Sequence 1, Appli
c  25    67.4    1.9     260   2   US-08-520-678A-29       Sequence 29, Appl
c  26    67.4    1.9     260   3   US-08-897-126-29        Sequence 29, Appl
   27    67.4    1.9     746   3   US-09-013-810-1         Sequence 1, Appli
   28    67.2    1.9    1445   4   US-09-814-951A-1        Sequence 1, Appli
   29    67.2    1.9    2320   4   US-09-202-904A-13       Sequence 13, Appl
   30    67      1.9    3275   4   US-09-370-838-151       Sequence 151, App
   31    66.8    1.9    2852   3   US-09-027-137-2         Sequence 2, Appli
   32    66.8    1.9    2852   4   US-09-344-441-2         Sequence 2, Appli
   33    66.8    1.9    3238   3   US-08-123-934A-5        Sequence 5, Appli
   34    66.8    1.9    3238   5   PCT-US94-10080-5        Sequence 5, Appli
   35    66.6    1.9    1507   4   US-09-453-323-1         Sequence 1, Appli
   36    66.6    1.9    3334   4   US-09-996-243-288       Sequence 288, App
   37    66.6    1.9    5173   1   US-08-242-677-1         Sequence 1, Appli
   38    66.4    1.9    1249   4   US-09-461-325-128       Sequence 128, App
c  39    66.4    1.9    1260   4   US-09-461-325-93        Sequence 93, Appl
   40    66.4    1.9    2665   4   US-08-971-089-5         Sequence 5, Appli
   41    66.4    1.9    2718   4   US-09-667-135-1         Sequence 1, Appli
   42    66.4    1.9    2773   4   US-09-996-243-178       Sequence 178, App
   43    66.2    1.9    1100   3   US-07-861-458C-4        Sequence 4, Appli
   44    66.2    1.9    2136   4   US-09-996-243-302       Sequence 302, App
   45    66.2    1.9    2218   4   US-09-016-434-1157      Sequence 1157, Ap
```

ALIGNMENTS

RESULT 1
US-09-904-615-66
; Sequence 66, Application US/09904615
; Patent No. 6566325
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: PZ032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0

```
  Query Match              2.1%;   Score 74;   DB 4;  Length 664;
  Best Local Similarity   64.0%;   Pred. No. 2.3e-08;
  Matches  110;  Conservative    1;  Mismatches   61;  Indels    0;  Gaps    0;


Qy      3297 CAAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCG 3356
                || |    |||    | |  | ||  ||    ||      |   |   |   ||| |||| ||
Db       487 CATAGTGTAAAAATTTATATTATTGTGAGGTTTTTTGTCTTTTTTTTTTTTTTTTTTTTT 546


Qy      3357 TGTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAA 3416
                || | |    ||| | ||   | |    |   | ||||||| |  ||   :    |   |||||||
Db       547 GGTATATTGCTGTATCTACTTTAACTTCCAGAAATAAACGTTATATRGGAAAAAAAAAAA 606


Qy      3417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       607 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 658
```

RESULT 2
US-09-328-965-1
; Sequence 1, Application US/09328965
; Patent No. 6501008
; GENERAL INFORMATION:
;   APPLICANT: Nevins, Donald J.
;   APPLICANT: Simmons, Carl
;   APPLICANT: The Regents of the University of California
;   TITLE OF INVENTION: Endo- and Exo-Glucanases and Gene
;   FILE REFERENCE: 023070-096600US
;   CURRENT APPLICATION NUMBER: US/09/328,965
;   CURRENT FILING DATE: 1999-06-09
;   EARLIER APPLICATION NUMBER: US 60/088,780
;   EARLIER FILING DATE: 1998-06-10
;   NUMBER OF SEQ ID NOS: 3
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1091
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: maize coleoptile endo-1,3;1,4-beta glucanase cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (68)..(979)
;   OTHER INFORMATION: endo-1,3;1,4-beta glucanase
US-09-328-965-1

```
Query Match              2.1%;   Score 73.8;  DB 4;   Length 1091;
Best Local Similarity   65.5%;   Pred. No. 3.2e-08;
Matches  108;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;


Qy       3304 TAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGT 3363
              ||||||  |   |  |   ||||  |           || |   || | || |||| | | | |
Db        918 TAATTTCCCTCATTTTTTTTGTCTCTATGTATTTCTTTTCTTTTCTTTTTGCTTTTTTAT 977


Qy       3364 GCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAA 3423
              |   |    | ||  |  |  |||||  |||   ||     |  |  |||||||||||||||
Db        978 GATCGCAATAAAGTTCAGTAGGGGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1037


Qy       3424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1038 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1082
```

RESULT 3
US-09-620-312D-706/c
; Sequence 706, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Liu, Chenghua
;   APPLICANT: Asundi, Vinod
;   APPLICANT: Zhang, Jie
;   APPLICANT: Ren, Feiyan
;   APPLICANT: Chen, Rui-hong
;   APPLICANT: Zhao, Qing A.
;   APPLICANT: Wehrman, Tom
;   APPLICANT: Xue, Aidong J.
;   APPLICANT: Yang, Yonghong
;   APPLICANT: Wang, Jian-Rui
;   APPLICANT: Zhou, Ping
;   APPLICANT: Ma, Yunqing
;   APPLICANT: Wang, Dunrui
;   APPLICANT: Wang, Zhiwei
;   APPLICANT: John Tillinghast
;   APPLICANT: Drmanac, Radoje T.
;   TITLE OF INVENTION: No. 6569662el Nucleic Acids and
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: 784CIP2B
;   CURRENT APPLICATION NUMBER: US/09/620,312D
;   CURRENT FILING DATE: 2000-07-19
;   PRIOR APPLICATION NUMBER: 09/552,317
;   PRIOR FILING DATE: 2000-04-25
;   PRIOR APPLICATION NUMBER: 09/488,725
;   PRIOR FILING DATE: 2000-01-21
;   NUMBER OF SEQ ID NOS: 1105
;   SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 706
;   LENGTH: 4055
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS

    Query Match   2.1%;   Score 73.4;   DB 4;   Length 4055;
    Best Local Similarity   58.4%; Pred. No. 6.9e-08;
    Matches 128; Conservative   0; Mismatches   91; Indels   0; Gaps   0;

Qy   3250 AAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTT 3309
             ||| | | | | ||||||| || | | | | || | | |||||
Db   500 AAAAGCAAAATGTGTTTTCAGATTTGTTACTTTAATAAAGGTTATCCATACCAATAAAAA 441

Qy   3310 GTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGT 3369
             || | | | | | | | | | |||| ||||| || | | ||| |
Db   440 GTGTACAACACAGCATTTTCTGTTAAATTATTATTGGTTTTCAGTTGTAATTTGGTATTT 381

Qy   3370 GTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAA 3429
             | | | |||| | |||||| | | ||||||||||||||||||||||||
Db   380 TTTCTGGCATGCGTTTATTAATTTATTAAATTGGCTTTTAGAAAAAAAAAAAAAAAAAAA 321

Qy   3430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             |||||||||||||||||||||||||||||||||||||||
Db   320 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 282

RESULT 4
US-09-996-243-114
; Sequence 114, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
;   APPLICANT: Ashkenazi,Avi J.
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Botstein,David
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Eaton,Dan L.
;   APPLICANT: Ferrara,Napoleone
;   APPLICANT: Fong,Sherman
;   APPLICANT: Gerber,Hanspeter
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Grimaldi,J.Christopher
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Kljavin,Ivar J.
;   APPLICANT: Napier,Mary A.
;   APPLICANT: Pan,James
;   APPLICANT: Paoni,Nicholas F.
;   APPLICANT: Roy, Margaret Ann
;   APPLICANT: Stewart, Timothy A.
;   APPLICANT: Tumas, Daniel
;   APPLICANT: Watanabe, Colin K.
;   APPLICANT: Williams, P. Mickey
;   APPLICANT: Wood, William I.
;   APPLICANT: Zhang, Zemin
;   TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;   TITLE OF INVENTION: Acids Encoding the Same
;   FILE REFERENCE: P2730P1C13
;   CURRENT APPLICATION NUMBER: US/09/996,243

; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10

```
;   PRIOR APPLICATION NUMBER: 60/088738
;   PRIOR FILING DATE: 1998-06-10
;   PRIOR APPLICATION NUMBER: 60/088742
;   PRIOR FILING DATE: 1998-06-10
;   PRIOR APPLICATION NUMBER: 60/088810
;   PRIOR FILING DATE: 1998-06-10
;   PRIOR APPLICATION NUMBER: 60/088824
;   PRIOR FILING DATE: 1998-06-10
;   PRIOR APPLICATION NUMBER: 60/088826
;   PRIOR FILING DATE: 1998-06-10
;   PRIOR APPLICATION NUMBER: 60/088858
;   PRIOR FILING DATE: 1998-06-11
;   PRIOR APPLICATION NUMBER: 60/088861
;   PRIOR FILING DATE: 1998-06-11
;   PRIOR APPLICATION NUMBER: 60/088876
;   PRIOR FILING DATE: 1998-06-11
;   PRIOR APPLICATION NUMBER: 60/089105
;   PRIOR FILING DATE: 1998-06-12
;   PRIOR APPLICATION NUMBER: 60/089440
;   PRIOR FILING DATE: 1998-06-16
;   PRIOR APPLICATION NUMBER: 60/089512
;   PRIOR FILING DATE: 1998-06-16
;   PRIOR APPLICATION NUMBER: 60/089514
;   PRIOR FILING DATE: 1998-06-16
;   PRIOR APPLICATION NUMBER: 60/089532
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 60/089538
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 60/089598
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 60/089599
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 60/089600
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 60/089653
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 60/089801
;   PRIOR FILING DATE: 1998-06-18
;   PRIOR APPLICATION NUMBER: 60/089907
;   PRIOR FILING DATE: 1998-06-18
;   PRIOR APPLICATION NUMBER: 60/089908
;   PRIOR FILING DATE: 1998-06-18
;   PRIOR APPLICATION NUMBER: 60/089947
;   PRIOR FILING DATE: 1998-06-19
;   PRIOR APPLICATION NUMBER: 60/089948
;   PRIOR FILING DATE: 1998-06-19
;   PRIOR APPLICATION NUMBER: 60/089952
;   PRIOR FILING DATE: 1998-06-19
;   PRIOR APPLICATION NUMBER: 60/090246
;   PRIOR FILING DATE: 1998-06-22
;   PRIOR APPLICATION NUMBER: 60/090252
;   PRIOR FILING DATE: 1998-06-22
;   PRIOR APPLICATION NUMBER: 60/090254
;   PRIOR FILING DATE: 1998-06-22
;   PRIOR APPLICATION NUMBER: 60/090349
;   PRIOR FILING DATE: 1998-06-23
;   PRIOR APPLICATION NUMBER: 60/090355
```

; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

     Query Match   2.1%; Score 71.2; DB 4; Length 1701;

```
     Best Local Similarity   73.4%;  Pred. No. 1.6e-07;
     Matches   91; Conservative   0; Mismatches  33; Indels   0; Gaps   0;

Qy      3345 TGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGA 3404
             | | ||| |    | | || |   || ||      |   ||| |||||||  | | |  |
Db      1563 TTTGTTACTTTTTCTTTGCTAATTTGGAAGATTAACTCATTTTTAATAAAATTATGTCTA 1622

Qy      3405 TGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3464
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1623 AGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1682

Qy      3465 AAAA 3468
             ||||
Db      1683 AAAA 1686


RESULT 5
US-08-821-994-63
; Sequence 63, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-63

  Query Match            2.0%;  Score 69.2;  DB 3;  Length 1441;
  Best Local Similarity   53.8%;  Pred. No. 4.5e-07;
  Matches  143; Conservative   0; Mismatches  123; Indels   0; Gaps   0;

Qy      3203 CTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTT 3262
             ||  | |  |||      |   ||| |  | | ||| | ||  ||  |      || |
Db      1169 CTCATGCAGTAATCAAATTGGGATTGTTATAAGTTAAATTAATCTTGTATTATTGTTTGT 1228

Qy      3263 ATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGT 3322
             || |  || ||| |  |    | || | |     |     | ||||  |||  |    |
Db      1229 ATGTATAGTATTTCGAAAAAAATTGATTCACCATAGGGATTTAATCTGTATAAATCTCTA 1288

Qy      3323 TGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATC 3382
             ||| | |||    | | ||      ||      | ||    |         |
Db      1289 TGTTGGTCAATATCATTTCATTCAAAGAATATTTGCTTTGGCTTGATTATGTATTAAGAG 1348

Qy      3383 ACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3442
```

```
         | |  |||||||    ||    |  | ||||||||||||||||||||||||||||||||
Db    1349 AAATATAATAAAAATGATATATTTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1408


Qy    3443 AAAAAAAAAAAAAAAAAAAAAAAAAA 3468
           ||||||||||||||||||||||||||
Db    1409 AAAAAAAAAAAAAAAAAAAAAAAAAA 1434
```

RESULT 6
US-09-363-708-3
; Sequence 3, Application US/09363708
; Patent No. 6399747
; GENERAL INFORMATION:
;   APPLICANT: Schmandt, et al.
;   TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 233 South Wacker Drive/6300 Sears Tower
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/363,708
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107
;   REFERENCE/DOCKET NUMBER: 01017/34451
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (312) 474-6300
;   TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2246 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   DESCRIPTION: /desc = "mouse PAL cDNA"
US-09-363-708-3

   Query Match   2.0%;   Score 69.2; DB 4;   Length 2246;
   Best Local Similarity   64.2%;   Pred. No. 5.4e-07;
   Matches 104; Conservative   0; Mismatches   58; Indels   0; Gaps   0;

```
Qy    3307 TTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCA 3366
           ||   ||  |  ||  | ||   |    |  |  |   ||  ||| | | ||  |||| |
Db    2054 TTGAACATGTCTTAAGTATGCTGCTTATATACTTTGCTTCATTTGCTTCATGGCTGTGTA 2113
```

```
Qy     3367  TGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAA  3426
             |      | |||||     |   || |||    ||      |  |  ||||||||||||||||||
Db     2114  TTATATAAAGTGTACTTGACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2173

Qy     3427  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3468
             ||||||||||||||||||||||||||||||||||||||||||
Db     2174  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2215
```

RESULT 7
US-09-083-587-3
; Sequence 3, Application US/09083587
; Patent No. 6492138
;   GENERAL INFORMATION:
;   APPLICANT: Schmandt, et al.
;   TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 233 South Wacker Drive/6300 Sears Tower
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/083,587
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clough, David W.
;   REGISTRATION NUMBER: 36,107
;   REFERENCE/DOCKET NUMBER: 01017/34451
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (312) 474-6300
;   TELEFAX: (312) 474-0448
;   INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2246 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   DESCRIPTION: /desc = "mouse PAL cDNA"
US-09-083-587-3

   Query Match   2.0%;   Score 69.2; DB 4;   Length 2246;
   Best Local Similarity 64.2%; Pred. No. 5.4e-07;
   Matches 104; Conservative   0; Mismatches 58; Indels   0; Gaps   0;

```
Qy     3307  TTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCA  3366
             ||   ||  |  ||  | ||   |    |   |  |  || |||  | | ||  |||| |
Db     2054  TTGAACATGTCTTAAGTATGCTGCTTATATACTTTGCTTCATTTGCTTCATGGCTGTGTA  2113
```

```
Qy    3367  TGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAA 3426
             |     | |||||    |   || |||   ||    |  |  ||||||||||||||||||
Db    2114  TTATATAAAGTGTACTTGACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2173

Qy    3427  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             ||||||||||||||||||||||||||||||||||||||||||
Db    2174  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2215
```

RESULT 8
US-09-594-506-37
; Sequence 37, Application US/09594506
; Patent No. 6512164
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: Thiamine Biosynthetic Enzymes
; FILE REFERENCE: BB1372 US NA
; CURRENT APPLICATION NUMBER: US/09/594,506
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/139,556
; PRIOR FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
;   LENGTH: 2406
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-594-506-37

```
   Query Match              2.0%;  Score 69.2;  DB 4;  Length 2406;
   Best Local Similarity   56.6%;  Pred. No. 5.6e-07;
   Matches  128; Conservative   0; Mismatches   98; Indels   0; Gaps   0;

Qy    3243  AAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAA 3302
             ||  | | ||| |||    || ||||  || | |||    | ||| ||       |   || |||
Db    2179  AACTTCATAAGCCCATAATTTTTTTGAGGAATCCCATTACATCTCGCAAAGCATTCACAA 2238

Qy    3303  ATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTG 3362
             |     |||| |    ||       | | |        |  ||  | |||       | |
Db    2239  TGTCCTGTGTAATTTACTTTTTACACCTATCCTTGTACATATTTCTATATAAGTAGAATA 2298

Qy    3363  TGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAA 3422
             | | |     |   | | | ||||  || |||   ||    |  | ||||||  |||||
Db    2299  TAAAAGATGTAACTAGATTGACAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCAAAAA 2358

Qy    3423  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             ||||||||||||||||||||||||||||||||||||||||||||||
Db    2359  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2404
```

RESULT 9
US-09-465-558-59
; Sequence 59, Application US/09465558
; Patent No. 6436657
; GENERAL INFORMATION:

```
  Query Match               2.0%;  Score 69;  DB 4;  Length 2202;
  Best Local Similarity   69.9%;  Pred. No. 6e-07;
  Matches   93;  Conservative    0;  Mismatches   40;  Indels   0;  Gaps   0;


Qy      3336 CTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAG 3395
             |||||  |||||||.|   ||| ||    |   | |||  |    | || |   ||| |
Db      2064 CTCCTTGTTTGTTTGCGTGCTTGGTGATCTGTATGAATGAAATAAATACGTGATTTAAGG 2123


Qy      3396 GCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3455
               ||    |  |   ||||||||||||||||||||||||||||||||||||||||||||||
Db      2124 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2183


Qy      3456 AAAAAAAAAAAAA 3468
             |||||||||||||
Db      2184 AAAAAAAAAAAAA 2196
```

; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1147 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 47..960
US-08-665-716-1

  Query Match   2.0%;   Score 68.8;   DB 1;   Length 1147;
  Best Local Similarity   75.9%;   Pred. No. 5.1e-07;
  Matches   85; Conservative   0; Mismatches   27; Indels   0; Gaps   0;

Qy   3357 TGTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAA 3416
             || || || ||| | | || |||| ||| | || | || |||||||
Db   1021 TGCGTAGACACATGTCCAAGGAGTTCTGGGGTATAAAAAGTTGTTCATTTATGAAAAAAA 1080

Qy   3417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1132

RESULT 11
US-09-182-816-22
; Sequence 22, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisnewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1736

```
   Query Match              2.0%;   Score 68.6;   DB 3;   Length 1736;
   Best Local Similarity   62.6%;   Pred. No. 6.8e-07;
   Matches  107;  Conservative     0;  Mismatches   64;  Indels   0;  Gaps    0;


Qy     3298 AAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGT 3357
            || |||| |||||| |      |     |   ||| |  ||  |    |||||  | | |
Db     1565 AATAAATTATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTACTGTGAAATAAA 1624


Qy     3358 GTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAA 3417
            |  || || |    |      |     |  || || || |||   ||   |  | | ||||||||
Db     1625 CGATATGGATTTTATTTCAAACTTGTCAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1684


Qy     3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1735
```


RESULT 12
US-09-182-816-24/c
; Sequence 24, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisnewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-182-816-24

```
   Query Match              2.0%;   Score 68.6;   DB 3;   Length 1736;
   Best Local Similarity   62.6%;   Pred. No. 6.8e-07;
   Matches  107;  Conservative     0;  Mismatches   64;  Indels   0;  Gaps    0;


Qy     3298 AAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGT 3357
            || |||| |||||| |      |     |   ||| |  ||  |    |||||  | | |
Db      172 AATAAATTATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTACTGTGAAATAAA 113


Qy     3358 GTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAA 3417
```

```
                 |  || || |     |       |  ||  ||  |||    ||    |  |  ||||||||
Db    112 CGATATGGATTTTATTTCAAACTTGTCAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 53


Qy   3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     52 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2
```

RESULT 13
US-09-471-528-22
; Sequence 22, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisnewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1736
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (159)..(1553)
US-09-471-528-22

   Query Match   2.0%;   Score 68.6;   DB 3;   Length 1736;
   Best Local Similarity   62.6%;   Pred. No. 6.8e-07;
   Matches 107; Conservative   0; Mismatches   64; Indels   0; Gaps   0;

```
Qy   3298 AAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGT 3357
          ||  |||| |||||| |       |     |   |||    ||  |      |||||  |  | |
Db   1565 AATAAATTATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTACTGTGAAATAAA 1624

Qy   3358 GTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAA 3417
            |  || || |   |        |    |  ||  ||  |||    ||     |  |  |  ||||||||
Db   1625 CGATATGGATTTTATTTCAAACTTGTCAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1684

Qy   3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1735
```

RESULT 14
US-09-471-528-24/c
; Sequence 24, Application US/09471528
; Patent No. 6153397

```
  Query Match              2.0%;   Score 68.6;   DB 3;   Length 1736;
  Best Local Similarity   62.6%;   Pred. No. 6.8e-07;
  Matches  107;  Conservative     0;  Mismatches   64;  Indels    0;  Gaps    0;


Qy      3298 AAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGT 3357
             || |||| ||||||| |      |     |   ||| |    || |     ||||| | | |
Db       172 AATAAATTATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTACTGTGAAATAAA 113


Qy      3358 GTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAA 3417
              | || || |     |     |    | || || ||| ||     | |    |||||||||
Db       112 CGATATGGATTTTATTTCAAACTTGTCAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAA 53


Qy      3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        52 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2
```

RESULT 15
US-09-634-530-22

```
   Query Match              2.0%;   Score 68.6;  DB 3;   Length 1736;
   Best Local Similarity   62.6%;   Pred. No. 6.8e-07;
   Matches  107;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy      3298 AAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGT 3357
             || |||| |||||| |     |     |   ||| || | | |||||   | | |
Db      1565 AATAAATTATTTGTGATAATAATATAATGTTAAAAATAAATGTAATTACTGTGAAATAAA 1624

Qy      3358 GTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAA 3417
              | || || |    |         |   | ||   || ||   ||| ||   | |  | ||||||||
Db      1625 CGATATGGATTTTATTTCAAACTTGTCAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAA 1684

Qy      3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1735
```

Search completed: January 29, 2004, 02:33:11
Job time : 195 secs

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 20:53:15 ; Search time 1084 Seconds
                                             (without alignments)
                                             11442.066 Million cell updates/sec

Title: US-10-056-884A-1
Perfect score: 3468
Sequence: 1 caagcactgtgctaaagtgt.........aaaaaaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2356869 seqs, 1788235258 residues

Total number of hits satisfying chosen parameters: 4713738

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Listing first 45 summaries

Database :   Published_Applications_NA:*
                   1:   /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                   2:   /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                   3:   /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                   4:   /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                   5:   /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                   6:   /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                   7:   /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                   8:   /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                   9:   /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                   10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                   11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                   12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                   13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
                   14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                   15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                   16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                   17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                   18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                SUMMARIES

		% Query				
Result No.	Score	Match	Length	DB	ID	Description
	1	3468	100.0	3468	15	US-10-056-884-1	Sequence 1, Appli
	2	769	22.2	769	15	US-10-056-884-3	Sequence 3, Appli
c	3	276.4	8.0	541	13	US-10-029-386-10927	Sequence 10927, A
c	4	240	6.9	279	13	US-10-029-386-24630	Sequence 24630, A
	5	205	5.9	632	15	US-10-060-036-4467	Sequence 4467, Ap
	6	201	5.8	614	15	US-10-060-036-564	Sequence 564, App
	7	167	4.8	2052	15	US-10-080-980-1	Sequence 1, Appli
c	8	156.2	4.5	978	13	US-10-029-386-20178	Sequence 20178, A
	9	104.6	3.0	688	15	US-10-080-980-8	Sequence 8, Appli
	10	87	2.5	249	11	US-09-918-995-2311	Sequence 2311, Ap
c	11	80.6	2.3	496	13	US-09-814-353-4862	Sequence 4862, Ap
c	12	80.6	2.3	496	13	US-09-814-353-11159	Sequence 11159, A
c	13	80	2.3	80	15	US-10-056-884-8	Sequence 8, Appli
c	14	79	2.3	425	10	US-09-834-975-451	Sequence 451, App
	15	79	2.3	1492	9	US-09-925-299-112	Sequence 112, App
	16	79	2.3	1492	11	US-09-925-299-112	Sequence 112, App
	17	79	2.3	1493	12	US-10-264-237-893	Sequence 893, App
	18	79	2.3	1493	15	US-10-106-698-1525	Sequence 1525, Ap
c	19	77	2.2	330	13	US-09-814-353-5205	Sequence 5205, Ap
c	20	77	2.2	330	13	US-09-814-353-11492	Sequence 11492, A
c	21	76.8	2.2	2796	12	US-10-264-237-1167	Sequence 1167, Ap
c	22	76.6	2.2	361	15	US-10-198-846-8619	Sequence 8619, Ap
c	23	76.4	2.2	388	13	US-09-814-353-17525	Sequence 17525, A
c	24	76.2	2.2	476	13	US-09-814-353-4700	Sequence 4700, Ap
c	25	76.2	2.2	476	13	US-09-814-353-10999	Sequence 10999, A
c	26	75.4	2.2	377	13	US-09-814-353-4680	Sequence 4680, Ap

```
c  27    75.4    2.2    377   13   US-09-814-353-10979      Sequence 10979, A
c  28    75.2    2.2    679   13   US-09-814-353-5129       Sequence 5129, Ap
c  29    75.2    2.2    679   13   US-09-814-353-11420      Sequence 11420, A
c  30      75    2.2    442    9   US-09-770-444-872      Sequence 872, App
c  31    74.6    2.2    392   13   US-09-814-353-17543      Sequence 17543, A
c  32    74.4    2.1    539   13   US-09-814-353-5852       Sequence 5852, Ap
c  33    74.4    2.1    539   13   US-09-814-353-12133      Sequence 12133, A
c  34    74.2    2.1    436   13   US-09-814-353-5048       Sequence 5048, Ap
c  35    74.2    2.1    436   13   US-09-814-353-11340      Sequence 11340, A
c  36      74    2.1    433   13   US-09-814-353-4798       Sequence 4798, Ap
c  37      74    2.1    433   13   US-09-814-353-11095      Sequence 11095, A
   38      74    2.1    664    9   US-09-739-254-66       Sequence 66, Appl
   39      74    2.1    664    9   US-09-904-615-66       Sequence 66, Appl
   40      74    2.1    664   13   US-10-055-098-66       Sequence 66, Appl
   41      74    2.1    664   15   US-10-054-988-66       Sequence 66, Appl
c  42    73.8    2.1    299   13   US-09-814-353-4844       Sequence 4844, Ap
c  43    73.8    2.1    299   13   US-09-814-353-11141      Sequence 11141, A
c  44    73.8    2.1    482   13   US-09-814-353-5268       Sequence 5268, Ap
c  45    73.8    2.1    482   13   US-09-814-353-11555      Sequence 11555, A
```

ALIGNMENTS

RESULT 1
US-10-056-884-1
; Sequence 1, Application US/10056884
; Publication No. US20030032786A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL
BETA-SUBUNIT,
;   TITLE OF INVENTION: K+betaM2
;   FILE REFERENCE: D0076 NP
;   CURRENT APPLICATION NUMBER: US/10/056,884
;   CURRENT FILING DATE: 2002-01-24
;   PRIOR APPLICATION NUMBER: US 60/263,872
;   PRIOR FILING DATE: 2001-01-24
;   PRIOR APPLICATION NUMBER: US 60/269,794
;   PRIOR FILING DATE: 2001-02-14
;   NUMBER OF SEQ ID NOS: 73
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 3468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (515)..(1798)
US-10-056-884-1

  Query Match   100.0%; Score 3468; DB 15; Length 3468;
  Best Local Similarity   100.0%; Pred. No. 0;
  Matches 3468; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CAAGCACTGTGCTAAAGTGTTTTTCATATGTCATGAAAAGTTGTGCCAGAAAATTATGGT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1   CAAGCACTGTGCTAAAGTGTTTTTCATATGTCATGAAAAGTTGTGCCAGAAAATTATGGT 60

Qy     61   TTGAACATGGGCAGTTTTCTCCTACCGTCAGCTATATCCACAAGCATCACATGAAGTGGA 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61   TTGAACATGGGCAGTTTTCTCCTACCGTCAGCTATATCCACAAGCATCACATGAAGTGGA 120

Qy    121   GATCTGGCAGCTCTGTGTATTTCAGTCAAGTTCCACAATGAAACCTGACAATAATGGTAA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121   GATCTGGCAGCTCTGTGTATTTCAGTCAAGTTCCACAATGAAACCTGACAATAATGGTAA 180

Qy    181   AAACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181   AAACCAATACGGACATCTGAGTAACTGGGGAATTGGCCTGCCTTGCATGTGAGCTTGATG 240

Qy    241   GAAGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCAC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241   GAAGATTGGATATAGACGAGTTGATTATATTTTATGAAGTAGCAGCTCACTACCATCCAC 300

Qy    301   CATCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301   CATCCAGGGTTTAAACTACTTTTTCAGCATCACTTCACCTGTGGACTCTTATACATTTTG 360

Qy    361   ATTTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361   ATTTCTTGGGGGAAAAATACTGGGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCT 420

Qy    421   TTTCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421   TTTCCCTTTCTTACAAGTTGATCCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTT 480

Qy    481   TAAATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481   TAAATCAAAATAGCAGCAGCAGAAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCG 540

Qy    541   TTATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541   TTATTATCCTCGAGAACAAGGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCT 600

Qy    601   GAATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601   GAATGTCGGGGGTCAAGTTTATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTC 660

Qy    661   CCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661   CCTCCTGTGGAAAATGTTTTCCCCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTC 720

Qy    721   CAAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721   CAAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCT 780

Qy    781   CAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGA 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781   CAGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGA 840

Qy    841   AGCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841   AGCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCA 900
```

```
Qy    901 AAGCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAG 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 AAGCCCAGATGAATTCTGCCACAGTGACTTTGAAGATGCCTCCCAAGGAAGCGACACAAG 960

Qy    961 AATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGG 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AATCTGCCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGG 1020

Qy   1021 TTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGT 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGT 1080

Qy   1081 TCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTT 1140

Qy   1141 GAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAA 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAA 1200

Qy   1201 ATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGC 1260

Qy   1261 CTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTC 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGGTC 1320

Qy   1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGA 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGA 1380

Qy   1381 TTGCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTGCTGCTGCAAGAATGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGA 1440

Qy   1441 CCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCAT 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCAT 1500

Qy   1501 CTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGG 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGG 1560

Qy   1561 CCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCT 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCT 1620

Qy   1621 GACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAGCTGTTAAAGAAAAGCT 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAGCTGTTAAAGAAAAGCT 1680

Qy   1681 CTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGA 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGA 1740
```

```
Qy    1741 TCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATA 1800

Qy    1801 AGGGAGGGCTGGGGCGGGGAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAA 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 AGGGAGGGCTGGGGCGGGGAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAA 1860

Qy    1861 AAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACA 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 AAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACA 1920

Qy    1921 ATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACA 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 ATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACA 1980

Qy    1981 GGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAATCTTTTTTAGTTATTTGTTTGTT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 GGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAATCTTTTTTAGTTATTTGTTTGTT 2040

Qy    2041 TACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGC 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAGCCAGCTACGTAAAAGTAGC 2100

Qy    2101 TGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCC 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTTCTCTCATCCTTCTACCTCC 2160

Qy    2161 CTCCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CTCCTTTGAATGAGGGTATGGTAGAAAAAGATCTGGCCCAATGGCATAAGTTTGGAATTT 2220

Qy    2221 TTAATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTT 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 TTAATTTTGGTTTTTCCTTTTGTTTATGGGGTTGGGGGGAATGGCAGATTTATATGACTT 2280

Qy    2281 TTCACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAA 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 TTCACTCAAATCTATATGTGCCAGTTTATATTGACTCCGTATGCATGAGTATTTGTGCAA 2340

Qy    2341 CACAAGCACAACTAAGTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTC 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CACAAGCACAACTAAGTATGTATATACACATGACGCACACGATGCCAGGGCCTAGACCTC 2400

Qy    2401 CCAAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCT 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 CCAAGGGCTGTGCTCCTGCTCCCAGCAGCCCTCTCTTAGAATATTTCAGATGGATGAGCT 2460

Qy    2461 TCTGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTC 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 TCTGACTCTTTCTTAAAATTCTTTTGGGAAGATTTCCCAGCCTTTCTTCACAACACTTTC 2520

Qy    2521 TAACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCT 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TAACATCAAATGACTCTCATCATCAACAAATTGTATTCCTTATTGTGAAATTAATACCCT 2580

Qy    2581 CAGGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGT 2640
```

```
                I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2581  CAGGCTCCATTTTACTGCTTTGCTCTTTGTCTGCATTAAGAGAGGATGAGGAGAGCTGGT  2640

Qy    2641  CAAACATTCCTTGTGTTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATG  2700
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2641  CAAACATTCCTTGTGTTAAAAAAATCAAACATTCATATCCACAAAATTTTCTGCTAAATG  2700

Qy    2701  ACTCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCA  2760
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2701  ACTCCACACTCAGCCTTCTCTACCCTGAACTGAATTATCACCCTTTTCTCCATGTTTTCA  2760

Qy    2761  GAGTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCT  2820
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2761  GAGTTCTTACTGCCCACAGTTTAATGGTGTGGCCTTTCCACATAATCCACATTAAGTTCT  2820

Qy    2821  GTGTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCC  2880
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2821  GTGTTCCTGTGTTGTTGTGGAACTAAGGACAACACACAGTACTTGAATAAGGGTCCGGCC  2880

Qy    2881  TTTTGTTTGTTTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAAT  2940
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2881  TTTTGTTTGTTTTAGAGAAAGTTGTATTCCACACACAACCTAATAATTTCTTATAAAAAT  2940

Qy    2941  TTTAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATT  3000
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    2941  TTTAAACTACAAAGCTACATTTTTACTTGCTTGTAGCCGTTTTTGTTTGCCTTTGGGATT  3000

Qy    3001  CGGGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACA  3060
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3001  CGGGCTTTGGCTGTGCCCATGCTAGGATTTAGCTGTGTCATTTTTATGATGTCTGTAACA  3060

Qy    3061  ACCCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCT  3120
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3061  ACCCAACAAGGTAACTGAAGCTCCAGAGTTAAGGTTTCAGATTTCTAAATGAAACTATCT  3120

Qy    3121  TTTTCAATTACATCCTGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCC  3180
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3121  TTTTCAATTACATCCTGACTTGTATAGACACAGCCAAAAAGAAACTGTTAATAGCCATCC  3180

Qy    3181  GTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACA  3240
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3181  GTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACA  3240

Qy    3241  AGAAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAA  3300
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3241  AGAAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAA  3300

Qy    3301  AAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTG  3360
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3301  AAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTG  3360

Qy    3361  TGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAA  3420
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
Db    3361  TGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAA  3420

Qy    3421  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3468
            I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
```

Db   3421 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468

RESULT 2
US-10-056-884-3
; Sequence 3, Application US/10056884
; Publication No. US20030032786A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL
BETA-SUBUNIT,
;   TITLE OF INVENTION: K+betaM2
;   FILE REFERENCE: D0076 NP
;   CURRENT APPLICATION NUMBER: US/10/056,884
;   CURRENT FILING DATE: 2002-01-24
;   PRIOR APPLICATION NUMBER: US 60/263,872
;   PRIOR FILING DATE: 2001-01-24
;   PRIOR APPLICATION NUMBER: US 60/269,794
;   PRIOR FILING DATE: 2001-02-14
;   NUMBER OF SEQ ID NOS: 73
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 769
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-056-884-3

     Query Match   22.2%; Score 769; DB 15; Length 769;
     Best Local Similarity 100.0%; Pred. No. 9.5e-172;
     Matches 769; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   393 AGGTCATTTTTTAATAAGTTAGCATCCTTTTCCCTTTCTTACAAGTTGATCCAAAGGATA 452
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGGTCATTTTTTAATAAGTTAGCATCCTTTTCCCTTTCTTACAAGTTGATCCAAAGGATA 60

Qy   453 AGGCTGTGACTCCATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAGAAGAAAGGGA 512
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGGCTGTGACTCCATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAGAAGAAAGGGA 120

Qy   513 CAATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTC 572
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGGTCCGCAGTTC 180

Qy   573 CCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCC 632
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTATTTTACTCGCC 240

Qy   633 ATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAG 692
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCCCCAAAGAGAG 300

Qy   693 ACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGAT 752
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGACAGAGATGGAT 360

Qy   753 TCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACT 812
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db        361 TCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTGCCTGATCACT 420

Qy        813 TTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCA 872
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCAGACTTGGTCA 480

Qy        873 AACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTG 932
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCACAGTGACTTTG 540

Qy        933 AAGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCG 992
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCTCCCTGCTCCCTGCCG 600

Qy        993 ACCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGG 1052
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGG 660

Qy       1053 GACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCT 1112
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCT 720

Qy       1113 TGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGA 1161
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGA 769


RESULT 3
US-10-029-386-10927/c
; Sequence 10927, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR GENE
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10927
;   LENGTH: 541
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC008716.6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.44
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.55
;   OTHER INFORMATION: NT HIT: AB037738.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: O53257, EVALUE 3.90e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AI345820.1, EVALUE 1.90e-01
US-10-029-386-10927

  Query Match            8.0%;  Score 276.4;  DB 13;  Length 541;
```

```
     Best Local Similarity    97.9%;   Pred. No. 4.8e-55;
     Matches   280;   Conservative    0;   Mismatches    6;   Indels    0;   Gaps    0;


Qy      1336 TGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 1395
                | ||||  |      |||||||||||||||||||||||||||||||||||||||||||||||||
Db       286 TTTCTTTTCAGGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 227


Qy      1396 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 1455
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       226 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 167


Qy      1456 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 1515
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       166 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 107


Qy      1516 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 1575
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       106 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 47


Qy      1576 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGATTCTG 1621
                ||||||||||||||||||||||||||||||||||||||||| ||||
Db        46 CCAACAGTCAGAGATGCGGCGGAAAAGCGACTTACTCCGGACTCTG 1
```

RESULT 4
US-10-029-386-24630/c
; Sequence 24630, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR GENE
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24630
;   LENGTH: 279
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC008716.6
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.44
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.55
;   OTHER INFORMATION: SWISSPROT HIT: P19836, EVALUE 2.30e+00
;   OTHER INFORMATION: NT HIT: AB037738.1, EVALUE 0.00e+00
US-10-029-386-24630

     Query Match   6.9%;   Score 240;   DB 13;   Length 279;
     Best Local Similarity   98.0%;   Pred. No. 1.4e-46;
     Matches   243;   Conservative   0;   Mismatches   5;   Indels   0;   Gaps   0;

```
Qy      1336 TGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 1395
```

```
          | |||| |      ||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 TTTCTTTTCAGGTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAA 189

Qy    1396 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 1455
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 TGGCAAAGGTGACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAG 129

Qy    1456 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 1515
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 CTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGAC 69

Qy    1516 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 1575
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 ACGCCAGACCAACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGAT 9

Qy    1576 CCAACAGT 1583
           ||||||||
Db       8 CCAACAGT 1
```

RESULT 5
US-10-060-036-4467
; Sequence 4467, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4467
;   LENGTH: 632
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 552, 569
;   OTHER INFORMATION: n = A,T,C or G
US-10-060-036-4467

   Query Match   5.9%; Score 205; DB 15; Length 632;
   Best Local Similarity   93.4%; Pred. No. 4.4e-38;
   Matches 214; Conservative   0; Mismatches   15; Indels   0; Gaps   0;

```
Qy    3206 AGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATT 3265
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15 AGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATT 74

Qy    3266 TTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGT 3325
```

```
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75  TTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGT  134


Qy    3326  GTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACA  3385
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135  GTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACA  194


Qy    3386  GGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAAA  3434
            |||||||||||||||||||||||||||      |  |  ||||  ||  |   ||
Db     195  GGTAATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA  243



RESULT 6
US-10-060-036-564
; Sequence 564, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Persing, David H.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Jiang, Yuqiu
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF PANCREATIC CANCER
;  FILE REFERENCE: 210121.566
;  CURRENT APPLICATION NUMBER: US/10/060,036
;  CURRENT FILING DATE:  2002-01-30
;  NUMBER OF SEQ ID NOS: 4560
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 564
;   LENGTH: 614
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 534, 551, 575, 576
;   OTHER INFORMATION: n = A,T,C or G
US-10-060-036-564

  Query Match              5.8%;   Score 201;  DB 15;   Length 614;
  Best Local Similarity   93.3%;   Pred. No. 3.9e-37;
  Matches  210;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;


Qy    3210  ACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATTTTCA  3269
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  ACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATTTTCA  60


Qy    3270  GCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGC  3329
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  GCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGC  120


Qy    3330  CAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTA  3389
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121  CAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTA  180


Qy    3390  ATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAAA  3434
```

```
            |||||||||||||||||||||          |  |  ||||  ||   |  ||
Db       181 ATAAAGGCAATTGGATGATATCTGTAGGAGGAAAACAATGACTAA 225
```

RESULT 7
US-10-080-980-1
; Sequence 1, Application US/10080980
; Publication No. US20030036115A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL
BETA-SUBUNIT,
; TITLE OF INVENTION: K+betaM6, EXPRESSED HIGHLY IN THE SMALL INTESTINE
; FILE REFERENCE: D0121 NP
; CURRENT APPLICATION NUMBER: US/10/080,980
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270,132
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/278,953
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 2052
;   TYPE: DNA
;   ORGANISM: homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (121)..(1095)
US-10-080-980-1

   Query Match   4.8%;   Score 167;   DB 15;   Length 2052;
   Best Local Similarity   64.6%;   Pred. No. 9.4e-29;
   Matches 267; Conservative   0; Mismatches 140; Indels   6; Gaps   1;

Qy   967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
                | | || ||| ||| | | ||| | ||| | ||| || | || ||| |
Db   705 CACGCCGTCCCAGTCGCTGGACGGCAGCCGGCGCTCGGGCTACATCACCATCGGCTACCG 764

Qy   1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
                || ||||| |||| | || | || | ||||| || |||||||| ||| |||| | ||
Db   765 CGGCTCCTACACCATCGGGCGGGACGCGCAGGCGGACGCCAAGTTCCGGCGAGTGGCGCG 824

Qy   1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
                || ||||| |||| || || |||| || || || |||| || || |||| ||
Db   825 CATCACCGTTTGCGGAAAGACGTCGCTGGCCAAGGAGGTGTTTGGGGACACCCTGAACGA 884

Qy   1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
                |||| | |||||| || || |||| || | ||||||||| | | ||| ||||| |||||
Db   885 AAGCCGGGACCCCGACCGTCCCCCGGAGCGCTACACCTCGCGCTATTACCTCAAGTTCAA 944

Qy   1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
                |||||| ||| || || || | | |||| |||| || ||||||||||||||| || |
Db   945 CTTCCTGGAGCAGGCCTTCGACAAGCTGTCCGAGTCGGGCTTCCACATGGTGGCGTGCAG 1004

Qy   1267 CTCATCGGTGACAGCATCTTT------CATCAACCAATATACAGATGACAAGATCTGGTC 1320
                ||| ||| || | || || || | | | | || ||||||||||||| |
```

```
Db        1005 CTCCACGGGCACCTGCGCCTTTGCCAGCAGCACCGACCAGAGCGAGGACAAGATCTGGAC 1064

Qy        1321 AAGCTACACTGAATATGTCTTCTACCGTGAGCCTTCCAGATGGTCACCCTCAC 1373
               ||||||||| || || ||||||| | | |||     |     | | |||||| |
Db        1065 CAGCTACACCGAGTACGTCTTCTGCAGGGAGTGAGCTCCCCAGACCCCTCGC 1117


RESULT 8
US-10-029-386-20178/c
; Sequence 20178, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR GENE
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20178
;   LENGTH: 978
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC000403.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  15
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 8.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.3
;   OTHER INFORMATION: SWISSPROT HIT: Q14681, EVALUE 2.00e-04
;   OTHER INFORMATION: EST_HUMAN HIT: BG387727.1, EVALUE 8.00e-64
;   OTHER INFORMATION: NT HIT: gi16163086, EVALUE 0.00e+00
US-10-029-386-20178

   Query Match             4.5%;   Score 156.2;  DB 13;  Length 978;
   Best Local Similarity   65.6%;  Pred. No. 2.2e-26;
   Matches  246;  Conservative    0;  Mismatches  123;  Indels    6;  Gaps    1;

Qy         967 CCCCCCTTCCTCCCTGCTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAG 1026
               | | || |||     ||| | | | ||    | ||| | ||| || | || ||| |
Db         376 CACGCCGTCCCAGTCGCTGGACGGCAGCCGGCGCTCGGGCTACATCACCATCGGCTACCG 317

Qy        1027 AGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCG 1086
               || |||| |||| | || | ||| | ||||| || |||||||| ||| |||| | ||
Db         316 CGGCTCCTACACCATCGGGCGGGACGCGCAGGCGGACGCCAAGTTCCGGCGAGTGGCGCG 257

Qy        1087 GATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGA 1146
               ||     ||||| |||| || || |||| || || || || ||||| || || |||| ||
Db         256 CATCACCGTTTGCGGAAAGACGTCGCTGGCCAAGGAGGTGTTTGGGGACACCCTGAACGA 197
```

```
Qy      1147 AAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAA 1206
             |||| | ||||| || ||  |||| ||  | ||||||||| | | ||| ||||| |||||
Db       196 AAGCCGGGACCCCGACCGTCCCCCGGAGCGCTACACCTCGCGCTATTACCTCAAGTTCAA 137

Qy      1207 GCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAA 1266
             ||||||     ||| || || | | |||| ||||  || |||||||||||||| || |
Db       136 CTTCCTGGAGCAGGCCTTCGACAAGCTGTCCGAGTCGGGCTTCCACATGGTGGCGTGCAG 77

Qy      1267 CTCATCGGTGACAGCATCTTT------CATCAACCAATATACAGATGACAAGATCTGGTC 1320
             ||| ||| ||    | ||      || || · | | | ||  || ||||||||||||| |
Db        76 CTCCACGGGCACCTGCGCCTTTGCCAGCAGCACCGACCAGAGCGAGGACAAGATCTGGAC 17

Qy      1321 AAGCTACACTGAATA 1335
             ||||||||| || ||
Db        16 CAGCTACACCGAGTA 2
```

RESULT 9
US-10-080-980-8
; Sequence 8, Application US/10080980
; Publication No. US20030036115A1
; GENERAL INFORMATION:
;   APPLICANT: Bristol-Myers Squibb Company
;   TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL
BETA-SUBUNIT,
;   TITLE OF INVENTION: K+betaM6, EXPRESSED HIGHLY IN THE SMALL INTESTINE
;   FILE REFERENCE: D0121 NP
;   CURRENT APPLICATION NUMBER: US/10/080,980
;   CURRENT FILING DATE: 2002-02-21
;   PRIOR APPLICATION NUMBER: US 60/270,132
;   PRIOR FILING DATE: 2001-02-21
;   PRIOR APPLICATION NUMBER: US 60/278,953
;   PRIOR FILING DATE: 2001-03-27
;   NUMBER OF SEQ ID NOS: 74
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 688
;   TYPE: DNA
;   ORGANISM: homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: wherein "N" is equal to "A", "C", "G" or "T".
US-10-080-980-8

  Query Match   3.0%; Score 104.6; DB 15; Length 688;
  Best Local Similarity   51.1%; Pred. No. 3e-14;
  Matches 192; Conservative   0; Mismatches 178; Indels   6; Gaps   1;

```
Qy       563 TCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGGTCAAGTTTAT 622
             ||||| |  ||        ||||||| ||  | || |||||||| || ||||| || || ||
Db       186 TCCGCGGAGCCACCGCTCTTCCCCGACATCGTGGAGCTGAACGTGGGGGGGCCAGGTGTAC 245

Qy       623 TTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCC 682
             | || || | | ||| || |     | || | || || || ||| |||| | |
Db       246 GTGACCCGGCGCTGCACGGTGGTGTCGGTGCCCGACTCGCTGCTCTGGCGCATGTTCACG 305
```

```
Qy     683 CCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGAC 742
           |  ||        | |  | || || |||  |||| ||| || | || ||  | |||
Db     306 CAGCAGCA------GCCGCAGGAGCTGGCCCGGGACAGCAAAGGCCGCTTCTTTCTGGAC 359

Qy     743 AGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTG 802
           | || || ||| | ||||| || || ||||| || || ||||| |||| | || |||
Db     360 CGGGACGGCTTCCTCTTCCGCTACATCCTGGATTACCTGCGGGACTTGCAGCTCGTGCTG 419

Qy     803 CCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCA 862
           || ||  |||| || ||     |  | ||| |  | || || || |||||| |||| |||
Db     420 CCCGACTACTTCCCCGAGCGCAGCCGGCTGCAGCGCGAGGCCGAGTACTTCGAGCTGCCA 479

Qy     863 GACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGCCCAGATGAATTCTGCCAC 922
           ||  |  ||
Db     480 GAGCTCGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 539

Qy     923 AGTGACTTTGAAGATG 938
              |   |   | |  |
Db     540 NNNNNNNTGCACAAGG 555


RESULT 10
US-09-918-995-2311
; Sequence 2311, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2311
;  LENGTH: 249
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-918-995-2311

  Query Match              2.5%;   Score 87;  DB 11;  Length 249;
  Best Local Similarity   65.2%;   Pred. No. 2.4e-10;
  Matches  161;  Conservative   0;  Mismatches   80;  Indels   6;  Gaps    2;

Qy    1236 CAGAGTGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTTCATCAACC 1295
           | |||    || |||||||||||||| ||||||| ||||||||| || || | ||| ||||||
Db       1 CCGAGGCCGGCTTCCACATGGTGGCGTGTAACTCCTCGGGCACCGCCGCCTTCGTCAACC 60

Qy    1296 AATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCGTGAGCCT- 1354
           | ||       || ||||||||||||    |||||||| || || | |||| |||    |||
Db      61 AGTACCGCGACGACAAGATCTGGAGCAGCTACACCGAGTACATTTTCTTCCGACCACCTC 120

Qy    1355 --TCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAG---GTGACA 1409
             | |  |||||   |||      || |      | |  ||| | ||| ||     |||| |
Db     121 AGAAAATAGTATCACCTAAACAAGAACATGAAGATAGGATACATGACCAAGTCACTGATA 180
```

```
Qy       1410 AAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCCAGT 1469
              ||| | | || || || || || || || ||||| |||| ||| || || || ||||||||| |
Db        181 AAGGAAGTGAAAGTGGGACTTCCTGGAATGAGCTCTTCACTTCCAGTTGGGACAGCCATT 240

Qy       1470 CTGAGGC 1476
              | ||||||
Db        241 CAGAGGC 247
```

RESULT 11
US-09-814-353-4862/c
; Sequence 4862, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4862
;   LENGTH: 496
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 156, 157, 160, 161, 162, 163, 164, 165, 167, 168, 169, 170,
;   LOCATION: 171, 172, 173, 174, 175, 196, 197, 198, 200, 203, 205, 206,
;   LOCATION: 219, 220, 228, 232, 240, 241, 244, 245, 246, 247, 249, 250,
;   LOCATION: 252, 253, 256, 258, 260, 263, 264, 265, 267, 268, 269
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 270, 271, 274, 275, 280, 287, 288, 289, 290, 303, 306, 317,
;   LOCATION: 318, 322, 331, 347, 348, 355, 361, 362, 364, 367, 368, 369,
;   LOCATION: 381, 383, 388, 393, 398, 404, 408, 409, 410, 411, 412, 413,
;   LOCATION: 414, 415, 416, 417, 418, 419, 420, 421, 423, 424, 435
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
```

    Query Match              2.3%;   Score 80.6;  DB 13;   Length 496;
    Best Local Similarity   47.3%;   Pred. No. 1.2e-08;
    Matches  131;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy      3192 TCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAA 3251
             | | | |||| |       | | |      || | |   | | | | |   ||    |||
Db       343 TTTTTTTTTTCCNCCCCTTTCNTTTNNAATTAAAAAANATNTTTTTTCCCCAANNNNNAAA 284

Qy      3252 AGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGT 3311
             | | |.     |        ||     | | | |        | |||||| |   || |
Db       283 AAANAAAANNAANNNNNTNNNTTNANTNTTNNTNNCNNNNGGNNAAAAAAANTTTNTTTT 224

Qy      3312 CAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGT 3371
             || |        | |      | ||| |||| || |   | |
Db       223 TTTNNAAAAAAAAAAAGNNANATNTNNNTTTTTTTTTTTTTTTTTTTNNNNNNNNNCNN 164

Qy      3372 GTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAA 3431
             |       | |   || |||  ||     | | |||||||||||||||||||||||||||
Db       163 NNNNGCNNCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 104

Qy      3432 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             |||||||||||||||||||||||||||||||||||||
Db       103 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 67


RESULT 12
US-09-814-353-11159/c
; Sequence 11159, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21

; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11159
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 156, 157, 160, 161, 162, 163, 164, 165, 167, 168, 169, 170,
; LOCATION: 171, 172, 173, 174, 175, 196, 197, 198, 200, 203, 205, 206,
; LOCATION: 219, 220, 228, 232, 240, 241, 244, 245, 246, 247, 249, 250,
; LOCATION: 252, 253, 256, 258, 260, 263, 264, 265, 267, 268, 269
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 270, 271, 274, 275, 280, 287, 288, 289, 290, 303, 306, 317,
; LOCATION: 318, 322, 331, 347, 348, 355, 361, 362, 364, 367, 368, 369,
; LOCATION: 381, 383, 388, 393, 398, 404, 408, 409, 410, 411, 412, 413,
; LOCATION: 414, 415, 416, 417, 418, 419, 420, 421, 423, 424, 435
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 436, 438, 441, 450, 451, 452, 453, 454, 455, 456, 457, 458,
; LOCATION: 470, 471, 472, 475, 477, 478, 481, 482, 484
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-11159

Query Match                2.3%;  Score 80.6;  DB 13;  Length 496;
Best Local Similarity    47.3%;  Pred. No. 1.2e-08;
Matches  131;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

```
Qy      3192 TCTGTATTTTACTAAGGTACCAATAGCTCTTTCATAGACTTGTGCTACAAGAAGGTTAAA 3251
             | | | |||| |      | | |    | |   || | |   | | | | |   ||    |||
Db       343 TTTTTTTTTTCCNCCCCTTTCNTTTNNAATTAAAAAANATNTTTTTTCCCCAANNNNAAA 284

Qy      3252 AGACCAGTTTTATTTTCAGCATTCCTCATGCATTTCAGTGGTAACCAAAAAATAATTTGT 3311
             | | |     |        || |  | | | | | |   ||||| |    || |
Db       283 AAANAAAANNAANNNNNTNNNTTNANTNTTNNTNNCNNNNGGNNAAAAAAANTTTNTTTT 224

Qy      3312 CAATTAATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGT 3371
               || |       | |          |  ||| |||| ||   | |
Db       223 TTTNNAAAAAAAAAAAGNNANATNTNNNTTTTTTTTTTTTTTTTTTTTNNNNNNNNNCNN 164

Qy      3372 GTATGTGTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAA 3431
             |         | |  || ||| || |     | | |||||||||||||||||||||||||
Db       163 NNNNGCNNCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 104

Qy      3432 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
             |||||||||||||||||||||||||||||||||||
Db       103 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 67
```

RESULT 13
US-10-056-884-8/c
; Sequence 8, Application US/10056884
; Publication No. US20030032786A1
; GENERAL INFORMATION:

```
  Query Match                2.3%;  Score 80;  DB 15;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-09;
  Matches   80; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy         783 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 842
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          80 GGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAG 21

Qy         843 CTGAATACTTCCAGCTCCCA 862
               ||||||||||||||||||||
Db          20 CTGAATACTTCCAGCTCCCA 1
```

```
   Query Match              2.3%;   Score 79;   DB 10;   Length 425;
   Best Local Similarity   50.8%;   Pred. No. 2.6e-08;
   Matches  184;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy     3107 AAATGAAACTATCTTTTTCAATTACATCCTGACTTGTATAGACACAGCCAAAAAGAAACT 3166
            ||| ||| || ||||| || | | || | | | | ||||| ||||
Db      381 AAAAAAAAAAATATTTTTTTTTTTTTTTTTTTTTTTCCAAAAAAAAAAAAAAAAAAACC 322

Qy     3167 GTTAATAGCCATCCGTCCATGTAACTCTGTATTTTACTAAGGTACCAATAGCTCTTTCAT 3226
            || |      |      | || | | | || |       || | | |
Db      321 CTTTTTTTTTTTTAAAAAAAGTTTTTTTTTTTAAAAACCCCCCCCTTTTTTTTTGGGGG 262

Qy     3227 AGACTTGTGCTACAAGAAGGTTAAAAGACCAGTTTTATTTTCAGCATTCCTCATGCATTT 3286
            |     | |        | ||| ||| ||| |||| || | |
Db      261 GGGGGGATTTTTTTTTTTTTGGAAAAACCCCCTTTTTTTTTTTTTTTTTAAAAAAAACG 202

Qy     3287 CAGTGGTAACCAAAAAATAATTTGTCAATTAATAGTTGTGTGCCAAGCACTCCTAATTTG 3346
            | || |  |||||| ||   |   || | || || | ||        | | |||
Db      201 GGGGGGGGAAAAAAAAAAAAACCCTTTTTTTTTTTTTTTGGGGGAAATTTTTTTTTTTT 142

Qy     3347 TTTTATTGCGTGTGTGTGCATGTGTGTATGTGTATCACAGGTAATAAAGGCAATTGGATG 3406
            |||| ||   |      |   | | | | | | || ||| || |
Db      141 TTTTTTTTTTTAAAATTTTTTTTTTTTTTTNNNAAAAAAAAAAAAAAAAAAAAAAAAAAA 82

Qy     3407 ATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3466
            |  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 22

Qy     3467 AA 3468
            ||
Db       21 AA 20
```

```
  Query Match              2.3%;  Score 79;  DB 9;  Length 1492;
  Best Local Similarity   70.2%;  Pred. No. 5.6e-08;
  Matches  106;  Conservative   0;  Mismatches  45;  Indels   0;  Gaps   0;


Qy       3318 ATAGTTGTGTGCCAAGCACTCCTAATTTGTTTTATTGCGTGTGTGTGCATGTGTGTATGT 3377
              | |  | | |   | ||| |  | | ||| |     |||| | || || ||  ||| |||
Db       1276 AGAAATATATTGGAGGCAAAGTTCAGTTGATGACAATTGTGTATATGTTACTGATGCTGT 1335


Qy       3378 GTATCACAGGTAATAAAGGCAATTGGATGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3437
                || |     |||||||| ||||| || || ||||||||||||||||||||||||||||
Db       1336 AAATTATTTTTAATAAAGAAAATTGTATTATCAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1395


Qy       3438 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
              |||||||||||||||||||||||||||||||
Db       1396 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1426
```

Search completed: January 29, 2004, 02:51:24
Job time : 1088 secs

OM nucleic - nucleic search, using sw model

Run on: January 28, 2004, 20:27:00 ; Search time 6858 Seconds
                                               (without alignments)
                                               12290.452 Million cell updates/sec

Title: US-10-056-884A-1
Perfect score: 3468
Sequence: 1 caagcactgtgctaaagtgt..........aaaaaaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
                   1:   em_estba:*
                   2:   em_esthum:*
                   3:   em_estin:*
                   4:   em_estmu:*
                   5:   em_estov:*
                   6:   em_estpl:*
                   7:   em_estro:*
                   8:   em_htc:*
                   9:   gb_est1:*
                   10: gb_est2:*
                   11: gb_htc:*
                   12: gb_est3:*
                   13: gb_est4:*
                   14: gb_est5:*
                   15: em_estfun:*
                   16: em_estom:*
                   17: em_gss_hum:*
                   18: em_gss_inv:*
                   19: em_gss_pln:*
                   20: em_gss_vrt:*
                   21: em_gss_fun:*
                   22: em_gss_mam:*
                   23: em_gss_mus:*
                   24: em_gss_pro:*
                   25: em_gss_rod:*
                   26: em_gss_phg:*
                   27: em_gss_vrl:*

```
        28:  gb_gss1:*
        29:  gb_gss2:*


        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                        SUMMARIES
                        %
        Result          Query
            No.   Score  Match Length DB    ID                Description
        ------------------------------------------------------------------------
    c    1    479   13.8    489  28   AQ536411           AQ536411 RPCI-11-3
         2    442.4 12.8    810  11   AK015313           AK015313 Mus muscu
    c    3    400.2 11.5    592  28   AQ525390           AQ525390 HS_5228_B
         4    352.2 10.2    784  14   CA463745           CA463745 AGENCOURT
         5    352   10.1    778  13   BU961910           BU961910 AGENCOURT
         6    350.6 10.1    952  14   BY714867           BY714867 BY714867
    c    7    338.6  9.8    489  10   BF391086           BF391086 UI-R-CA1-
         8    329    9.5   2332  11   AK043351           AK043351 Mus muscu
         9    327.4  9.4   2343  11   AK047519           AK047519 Mus muscu
        10    324.2  9.3   2584  11   AK045439           AK045439 Mus muscu
        11    322.8  9.3   2555  11   AK042569           AK042569 Mus muscu
        12    310    8.9    973  13   BQ713664           BQ713664 AGENCOURT
        13    305.4  8.8    319   9   AA332022           AA332022 EST35911
        14    286.2  8.3    424  14   BY706433           BY706433 BY706433
        15    286    8.2    422  11   AK006368           AK006368 Mus muscu
    c   16    284.2  8.2    419  10   BF413617           BF413617 UI-R-CA0-
        17    282.6  8.1   1694  11   AK039167           AK039167 Mus muscu
        18    280.6  8.1    401  11   AK005863           AK005863 Mus muscu
        19    277    8.0    943  13   BU938472           BU938472 AGENCOURT
        20    270.2  7.8    400  14   BY706005           BY706005 BY706005
    c   21    257.6  7.4    386  10   BF413618           BF413618 UI-R-CA0-
        22    255.4  7.4   1710  11   AK046557           AK046557 Mus muscu
        23    255.4  7.4   1898  11   AK050097           AK050097 Mus muscu
        24    255.4  7.4   3552  11   AK085035           AK085035 Mus muscu
        25    255.4  7.4   3572  11   AK046708           AK046708 Mus muscu
        26    241.2  7.0    579  12   BJ095683           BJ095683 BJ095683
        27    241.2  7.0    584  12   BJ096172           BJ096172 BJ096172
        28    234.8  6.8    491  12   BJ095666           BJ095666 BJ095666
        29    228.4  6.6    870  14   CA470866           CA470866 AGENCOURT
        30    220.2  6.3    675  10   BB653208           BB653208 BB653208
        31    220    6.3   1746  11   AK082563           AK082563 Mus muscu
        32    219.2  6.3    295  10   BF522617           BF522617 UI-R-G0-u
        33    217.6  6.3    680  10   BB183090           BB183090 BB183090
    c   34    217.4  6.3    324   9   AI596442           AI596442 vi69b03.x
        35    207.2  6.0    433   9   AA497980           AA497980 vi69f03.r
        36    201    5.8    638  12   BM947722           BM947722 UI-M-EG0p
        37    190.6  5.5    927  14   CA964920           CA964920 CcLX05a17
        38    188.8  5.4    709  13   BU294624           BU294624 603604677
        39    188.4  5.4    304   9   AV046252           AV046252 AV046252
        40    186.4  5.4    594   9   AL588086           AL588086 AL588086
        41    179.2  5.2    310   9   AV206002           AV206002 AV206002
        42    170.8  4.9    324   9   AV206367           AV206367 AV206367
        43    169    4.9    295   9   AV040180           AV040180 AV040180
        44    167.6  4.8    642  12   BJ031784           BJ031784 BJ031784
        45    167    4.8   1003  29   CNS05PKF           AL348072 Tetraodon
```

ALIGNMENTS

RESULT 1
AQ536411/c
LOCUS   AQ536411   489 bp   DNA   linear   GSS 18-MAY-1999
DEFINITION   RPCI-11-318B21.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-318B21
             , genomic survey sequence.
ACCESSION   AQ536411
VERSION   AQ536411.1 GI:4848101
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 489)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
             ,J.C.
  TITLE   Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
             Map Building
  JOURNAL   Unpublished
  COMMENT   Other_GSSs: RPCI-11-318B21.TV
             Contact: Shaying Zhao, William Nierman, Mark Adams
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: hbe@tigr.org
             Clones are derived from the human BAC library RPCI-11. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
             Research Genet cs (info@resgen.com). BAC end search page:
             http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
             Seq primer: SP6
             Class: BAC ends.
FEATURES   Location/Qualifiers
     source   1..489
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7621772"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-318B21"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT   131 a   114 c   105 g   137 t   2 others
ORIGIN

  Query Match   13.8%; Score 479; DB 28; Length 489;
  Best Local Similarity   98.6%; Pred. No. 3.3e-42;
  Matches 482; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

```
Qy    383 GGATAAGAGGAGGTCATTTTTTAATAAGTTAGCATCCTTTTCCCTTTCTTACAAGTTGAT 442
          |||||||||  ||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    489 GGATAAGAGNAGGTCATTTTTTAATAAGTTAGCATCCTTTTCCCTATCTTACAAGTTGAT 430

Qy    443 CCAAAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAG 502
          |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 CCANAGGATAAGGCTGTGACTCCATTGGATTGCACCTTTAAATCAAAATAGCAGCAGCAG 370

Qy    503 AAGAAAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGG 562
          ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 AAGACAGGGACAATGGCTCTGAGTGGAAACTGTAGTCGTTATTATCCTCGAGAACAAGGG 310

Qy    563 TCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTAT 622
          |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db    309 TCCGCAGTTCCCAACTCCTTCCGTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTTTAT 250

Qy    623 TTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCC 682
          |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 TTTACTCGCCATACCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTTTCC 190

Qy    683 CCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGAC 742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 CCAAAGAGAGACACGGCTAATGATCTAGCCAAGGACTCCAAGGGAAGGTTTTTCATTGAC 130

Qy    743 AGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTG 802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 AGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGGGACAGGCAGGTGGTCCTG 70

Qy    803 CCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCCA 862
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     69 CCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCTGAATACTTCCAGCTCCAA 10

Qy    863 GACTTGGTC 871
          |||||||||
Db      9 GACTTGGTC 1
```

RESULT 2
AK015313

REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
    TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
    TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
    TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 810)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,

Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

```
  TITLE      Direct Submission
  JOURNAL    Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
  COMMENT    Please visit our web site (http://genome.gsc.riken.go.jp/) for
             further details.
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues. First strand cDNA was primed with a primer
             [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse transcriptase
             and subsequently enriched for full-length by cap-trapper. Second
             strand cDNA was prepared with the primer adapter of sequence [5'
             GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
             with BamHI and XhoI. cDNA of size comprised longer than 7 kb was
             selected before cloning. Vector: a modified pBluescript KS(+) after
             bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
             end: BamHI. Host: DH10B.
  FEATURES             Location/Qualifiers
       source          1. .810
                       /organism="Mus musculus"
                       /mol_type="mRNA"
                       /strain="C57BL/6J"
                       /db_xref="FANTOM_DB:4930434H12"
                       /db_xref="MGI:1896697"
                       /db_xref="taxon:10090"
                       /clone="4930434H12"
                       /sex="male"
                       /tissue_type="testis"
                       /clone_lib="RIKEN full-length enriched mouse cDNA library"
                       /dev_stage="adult"
       misc_feature    1. .810
                       /note="inferred: RIKEN cDNA 4930434H12 gene / putative
                       [Mus musculus] (UniGene|Mm.46143, TIGR-MGI1|TC1870,
                       evidence: UG/TGI)"
                       /db_xref="MGI:1914659"
  BASE COUNT      226 a    192 c    208 g    184 t
  ORIGIN
```

```
  Query Match             12.8%;  Score 442.4;  DB 11;  Length 810;
  Best Local Similarity   83.9%;  Pred. No. 2e-38;
  Matches  554; Conservative   0; Mismatches   86; Indels   20; Gaps    4;
```

Qy   1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406

```
                 |||||||||||| | ||||| |||| || || || |||||||||||||||||||||| || |
Db       167 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG 226

Qy      1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
                 |||   |||| ||||||||||| || |||||||||||| ||||| || ||||| |||||||
Db       227 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC 283

Qy      1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
                 |||| ||||||||||||||| |||||||||| |||||||| || || || ||||||| ||
Db       284 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 343

Qy      1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
                 |||||||||||||||| || |||||||||| || || || |||||||||||||||||||||
Db       344 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 403

Qy      1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
                 ||||| ||||||||| ||| | |||||||| ||||| ||||||||||||||| |||||||
Db       404 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA 463

Qy      1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
                 | ||||||||||| ||||||| || ||||||||||| || ||||| ||||||||||||||
Db       464 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT 523

Qy      1707 GTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
                 |||||||||||||| | || | |||||||||||| || |||||| |||||||||||||||
Db       524 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAAACATCCTT 583

Qy      1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
                 |||| ||||||||||| |||||||||||||||||||| ||| || | ||
Db       584 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGAGGGCTGTGG---------- 632

Qy      1827 AAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
                 |   | || ||||||||| |||||| |||||  ||||| |||| ||||||||||||||||
Db       633 ----GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAGACATATGTTAAAGGAAAAA 688

Qy      1887 AATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTA 1946
                 | ||||||||| | | ||||||| ||||| |||||| |||||| |||||||||| |||||
Db       689 TA-ACAACTAACGGTCCACATTTGTTAGATCACAAT-GTCCATTGATGTACTACTGCCTA 746

Qy      1947 CTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA 2006
                 |||| ||||| |||||||||| |||||||||||||||||||||||||||||||||||||||
Db       747 CTTTGCCTAGCTCACCTTAACGTGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA 806


RESULT 3
AQ525390/c
LOCUS       AQ525390                592 bp    DNA     linear   GSS 11-MAY-1999
DEFINITION  HS_5228_B2_C05_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=804 Col=10 Row=F, genomic survey sequence.
ACCESSION   AQ525390
VERSION     AQ525390.1  GI:4772710
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 592)
```

COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 804   row: F   column: 10
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 592.
FEATURES   Location/Qualifiers
     source   1. .592
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=804 Col=10 Row=F"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT   157 a   139 c   133 g   158 t   5 others
ORIGIN

   Query Match   11.5%;   Score 400.2; DB 28;   Length 592;
   Best Local Similarity   96.2%;   Pred. No. 7.6e-34;
   Matches   408; Conservative   0; Mismatches   16; Indels   0; Gaps   0;

Qy   928 CTTTGAAGATGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCCTTCCTCCCTGCTCCC 987
             ||||||||||||||||||||||||| |||||||||| |||| || ||||||| ||||||
Db   545 CTTTGAAGATGCCTCCCAAGGAAGAGACACAAGAATGTGCCNCCTTTCCTCCGTGCTCCA 486

Qy   988 TGCCGACCGCAAGTGGGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACCTTGGGCAG 1047
             ||| |||||||||| ||||||||||||||||||||||||||||||||||| ||||||||
Db   485 TGCGGACCGCAAGTGAGGTTTCATTACTGTGGGTTACAGAGGATCCTGCACTTTGGGCAG 426

Qy   1048 AGAGGGACAGGCAGATGCCAAGTTTCGGAGAGTTCCCCGGATTTTGGTTTGTGGAAGGAT 1107
             |||||||||||||||||||||||||| ||||||||||||| ||| |||||||||||||||
Db   425 AGAGGGACAGGCAGATGCCAAGTTTCGNAGAGTTCCCCGGATTNTGGTTTGTGGAAGGAT 366

```
Qy      1108 TTCCTTGGCAAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGC 1167
             |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       365 TTCCTTGGCGAAAGAAGTCTTTGGAGAAACTTTGAATGAAAGCAGAGACCCTGATCGAGC 306

Qy      1168 CCCAGAAAGATACACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGA 1227
             |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       305 GCCAGATAGATACACCTCCAGATTTTATCTCAAATTCAAGCACCTGGAAAGGGCTTTTGA 246

Qy      1228 TATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTT 1287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       245 TATGTTGTCAGAGTGTGGATTCCACATGGTGGCCTGTAACTCATCGGTGACAGCATCTTT 186

Qy      1288 CATCAACCAATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCG 1347
             |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       185 CATCATCCAATATACAGATGACAAGATCTGGTCAAGCTACACTGAATATGTCTTCTACCG 126

Qy      1348 TGAG 1351
             | ||
Db       125 TAAG 122


RESULT 4
CA463745
LOCUS       CA463745                784 bp    mRNA    linear   EST 12-NOV-2002
DEFINITION  AGENCOURT_10724816 NIH_MGC_169 Mus musculus cDNA clone
            IMAGE:6771233 5', mRNA sequence.
ACCESSION   CA463745
VERSION     CA463745.1  GI:24920097
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jonathan Kuo, NIMH
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3090  row: h  column: 16
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6771233"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_169"
                     /note="Organ: Testicles; Vector: pDNR-LIB; Site_1: SfiI
```

(ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in thelaboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

BASE COUNT   244 a   239 c   175 g   121 t   5 others
ORIGIN

    Query Match   10.2%;   Score 352.2; DB 14; Length 784;
    Best Local Similarity 85.9%; Pred. No. 8.5e-29;
    Matches 403; Conservative   0; Mismatches 63; Indels   3; Gaps   1;

Qy   1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
             |||||||||| | ||||| |||| || || || |||||||||||||||||||| || |
Db   104 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG 163

Qy   1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
             ||| |||| ||||||||||| || ||||||||||| ||||| || ||||| |||||||
Db   164 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC 220

Qy   1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| ||||||||||||||| |||||||||||| |||||||| || || || ||||||| ||
Db   221 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 280

Qy   1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
             ||||||||||||||||| || |||||||||| || || || ||||||||||||||||||||
Db   281 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 340

Qy   1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
             ||||| |||||||||.|| ||| | ||||||| |||||| |||||||||||||| |||||||
Db   341 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA 400

Qy   1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
             | ||||||||||| ||||||| || ||||||||||| || ||||| |||||||||||||||
Db   401 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT 460

Qy   1707 GTATCCAGGATTTCCTAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
             ||||||||||||||| | || | ||||||||||||||| || |||||| ||||||||||||
Db   461 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAAACATCCTT 520

Qy   1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGG 1815
             |||| |||||||||||| || |||||||||||| ||| | | | ||
Db   521 GGCAGTCTGAACTTTTACGGGAGTATCATCTATAGGGGGGAGGCTGTGG 569

RESULT 5
BU961910
LOCUS   BU961910   778 bp   mRNA   linear   EST 21-OCT-2002
DEFINITION AGENCOURT_10617166 NIH_MGC_169 Mus musculus cDNA clone
            IMAGE:6742567 5', mRNA sequence.
ACCESSION   BU961910
VERSION   BU961910.1 GI:24191482
KEYWORDS   EST.

```
SOURCE         Mus musculus (house mouse)
  ORGANISM     Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1  (bases 1 to 778)
  AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
  TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL      Unpublished
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Dr. Jonathan Kuo, NIMH
                cDNA Library Preparation: Michael Brownstein Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Agencourt Bioscience Corporation
                Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM3080  row: n  column: 06
               High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6742567"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_169"
                     /note="Organ: Testicles; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.5 kb
                     size fraction. Library created in thelaboratory of M.
                     Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
BASE COUNT        226 a      180 c      199 g      160 t       13 others
ORIGIN

   Query Match                10.1%;   Score 352;   DB 13;   Length 778;
   Best Local Similarity     85.6%;   Pred. No. 8.9e-29;
   Matches   404;  Conservative     0;  Mismatches    65;  Indels    3;  Gaps    1;


Qy      1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
             ||||||||||| | ||||| |||| || || || |||||||||||||||||||| || |
Db       183 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG 242

Qy      1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
             |||     |||| |||||||||||| || ||||||||||| ||||| || ||||| |||||||
Db       243 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC 299

Qy      1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| |||||||||||||||| |||||||||||| ||||||||| || || || ||||||| ||
Db       300 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 359

Qy      1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
```

```
          |||||||||||||||||| || |||||||||||| || || || |||||||||||||||||||||
Db    360 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 419

Qy   1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
          ||||| |||||||||| ||| | ||||||| |||||| ||||||||||||||| |||||||
Db    420 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA 479

Qy   1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
          | |||||||||||| ||||||| || |||||||||| || ||||| |||||||||||||||
Db    480 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT 539

Qy   1707 GTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
          |||||||||||||| | || | |||||||||||||||| || ||||| |||||||||||||
Db    540 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAACGAAAACATCCTT 599

Qy   1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGG 1818
          |||| ||||||||||||| |||||||||||||||||| ||| | |    | | ||
Db    600 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGGGAGGGCTGTGGGG 651
```

RESULT 6
BY714867

```
LOCUS       BY714867                 952 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY714867 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 4930434H12 5', mRNA sequence.
ACCESSION   BY714867
VERSION     BY714867.1  GI:27127984
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 952)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
```

                    ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
                    ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
                    of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
  COMMENT   Contact: Yoshihide Hayashizaki
                    Laboratory for Genome Exploration Research Group, RIKEN Genomic
                    Sciences Center(GSC), Yokohama Institute
                    The Institute of Physical and Chemical Research (RIKEN)
                    1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                    Tel: 81-45-503-9222
                    Fax: 81-45-503-9216
                    Email: genome-res@gsc.riken.go.jp,
                    URL:http://genome.gsc.riken.go.jp/
                    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
                    ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
                    Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
                    ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
                    Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
                    Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
                    Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
                    Submission
                     Computational Analysis of Full-Length Mouse cDNAs Compared with
                    Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
                     Normalization and subtraction of cap-trapper-selected cDNAs to
                    prepare full-length cDNA libraries for rapid discovery of new
                    genes. Genome Res. 10 (10), 1617-1630 (2000)
                     RIKEN integrated sequence analysis (RISA) system--384-format
                    sequencing pipeline with 384 multicapillary sequencer. Genome Res.
                    10 (11), 1757-1771 (2000)
                     Computer-based methods for the mouse full-length cDNA
                    encyclopedia: real-time sequence clustering for construction of a
                    nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
                     cDNA library was prepared and sequenced in Mouse Genome
                    Encyclopedia Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                    Division of Experimental Animal Research in Riken contributed to
                    prepare mouse tissues.
                     Please visit our web site (http://genome.gsc.riken.go.jp) for
                    further details.
  FEATURES   Location/Qualifiers
     source   1. .952
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="4930434H12"
                    /sex="male"
                    /tissue_type="testis"
                    /dev_stage="adult"
                    /lab_host="SOLR"
                    /clone_lib="RIKEN full-length enriched, adult male testis"
                    /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites. "
BASE COUNT   233 a   259 c   226 g   229 t   5 others
ORIGIN

  Query Match   10.1%;   Score 350.6; DB 14;   Length 952;
  Best Local Similarity   75.5%;   Pred. No. 1.1e-28;
  Matches 506; Conservative   0; Mismatches 142; Indels   22; Gaps   5;

Qy   1347 GTGAGCCTTCCAGATGGTCACCCTCACACTGCGATTGCTGCTGCAAGAATGGCAAAGGTG 1406
             |||||||||||| | ||||| |||| || || || ||||||||||||||||||||| || |
Db   167 GTGAGCCTTCCCGGTGGTCCTCCTCTCATTGTGACTGCTGCTGCAAGAATGGCAAGGGAG 226

Qy   1407 ACAAAGAAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCC 1466
             ||| |||| |||||||||||| || ||||||||||| ||||| || ||||| |||||||
Db   227 ACA---AAGGAGAGAGCGGCACCTCCTGCAATGACCTGTCCACTTCCAGCTGTGACAGCC 283

Qy   1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| |||||||||||||| ||||||||||| |||||||| || || || ||||||| ||
Db   284 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 343

Qy   1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
             |||||||||||||||| || ||||||||||| || || || ||||||||||||||||||||
Db   344 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 403

Qy   1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
             ||||| |||||||||| ||| | |||||||| |||||| |||||||||| |||| |||||||
Db   404 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCCGGGAGTCGAACA 463

Qy   1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
             | |||||||||||| ||||||| || |||||||||| || || || |||||||||||||
Db   464 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGACGAAGAGCTGGAGAAAT 523

Qy   1707 GTATCCAGGATTTCCTAAAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
             |||||||||||||| | || | |||||||||||||| || |||||| || |||||||
Db   524 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGATCACATCCTT 583

Qy   1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
             |||| ||||||||| ||| || |||||||||||||| ||| ||| | || |
Db   584 GGCAGTCTGAACTCTTACGGGAGTATCATCTATAGGGGGAGGCGTGCGGGCG-------- 635

Qy   1827 AAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
                     | | | |||| ||| | ||||| ||||| |||| | |||
Db   636 -------CTCGCGCTCTCAAATACACCCNCTCAAAGGGCGACATATCCTAAACG-GAAA 686

Qy   1887 AATACAACTAATGATGCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTA 1946
             | || |||||| |||||| | | | | |||| || ||| || ||||| ||||
Db   687 AGTATCACTAATCGNTCACATCTGTCACAGCACACACGT-CATCTATGTACTATCTCCTA 745

```
Qy      1947 CTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAA 2006
             ||||  |||  | ||||||| || | ||    ||      ||| ||| || ||    |
Db       746 CTTTGNCTA-CCCTCCTTAACGTGCCACTCACAGGGCACACTTTTTTTATATGTGGATCA 804

Qy      2007 GTACAAGAAA 2016
             || || | |
Db       805 CTATAATATA 814
```

RESULT 7
BF391086/c

```
LOCUS       BF391086                489 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  UI-R-CA1-bcd-a-05-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bcd-a-05-0-UI 3', mRNA sequence.
ACCESSION   BF391086
VERSION     BF391086.1  GI:11375933
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized testis library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com) The following repetitive elements were
            found in this cDNA sequence: 1-35, >POLY_A#Simple_repeat
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..489
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA1-bcd-a-05-0-UI"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-CA1"
```

```
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex, corpus striatum, testis, and
                     hippocampus. For a detailed description of the library
                     from which this clone was derived, please visit our web
                     site at ratest.eng.uiowa.edu. The subtraction has been
                     previously described in (Bonaldo, Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_LIB=UI-R-CA1
                     TAG_TISSUE=testis
                     TAG_SEQ=ACGCAG"
BASE COUNT        89 a      123 c      111 g      164 t      2 others
ORIGIN
```

```
    Query Match                9.8%;  Score 338.6;  DB 10;  Length 489;
    Best Local Similarity    84.0%;  Pred. No. 3.1e-27;
    Matches  400; Conservative    0; Mismatches   60; Indels   16; Gaps    1;


Qy      1413 AAGGGGAGAGCGGCACGTCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCCAGTCTG 1472
             |||||||||| ||||| || |||||||||||||| || || |||||||||||||||||| |
Db       483 AAGGGGAGAGTGGCACTTCCTGCAATGACCTCTCTACTTCCAGCTGCGACAGCCAGTCAG 424


Qy      1473 AGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCAACATCC 1532
             |||||||||||||||||||||||| || |||||||| || || || || ||| |||||||
Db       423 AGGCCAGCTCTCCCCAGGAGACAGTGATCTGTGGGCCTGTAACGCGTCAGGGCAACATCC 364


Qy      1533 AGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAGAGATGC 1592
             |||||||||||| ||||||||||| |||||||| || |||||||||||||||||||||||
Db       363 AGACTCTGGACCGGCCCATCAAGAAAGGCCCCGTGCAGCTGATCCAACAGTCAGAGATGA 304


Qy      1593 GGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACATGAGCA 1652
             |||||||||| ||| | |||||||| |||||||||| ||||| ||||| ||||||||| ||||
Db       303 GGCGGAAAAGTGACCTGCTCCGGACTCTGACTTCCGGCTCTAGGGAGTCGAACATNAGCA 244


Qy      1653 GCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAATGTATCC 1712
             ||||||| ||||||| || ||||||| |||| || ||||| |||||||||||||||||||||
Db       243 GCAAAAAGAAAGCTGCGAAGGAAAAGGTCTCCATCGAGGAAGAGCTGGAGAAATGTATCC 184


Qy      1713 AGGATTTCCTAAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTTGGCAAT 1772
             |||||||||| || | |||||||||||||| || |||||||||||||||||||||||||| |
Db       183 AGGATTTCCTGAAGATAAAAATTCCAGATCGCTTCCCTGAGAGAAAACATCCTTGGCAGT 124


Qy      1773 CTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAAAAAAAA 1832
             |||||||||| ||||||||||||||||| ||||||| | ||| | |
Db       123 CTGAACTTTTACGGAAGTATCATCTATAGGGGAGGGCAGTGGGTAGTCA---------- 75


Qy      1833 AAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAAAA 1888
             || ||||||||| |||||| | |||||||| ||||| ||||||||||||| |
Db        74 -----CCACTTTGAAATAAACCTCCTGAAAGGAAGACATATATTAAAGGAAAAATA 24
```

RESULT 8
AK043351
LOCUS   AK043351   2332 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length

```
                    enriched library, clone:A730087N02 product:hypothetical protein,
                    full insert sequence.
ACCESSION           AK043351
VERSION             AK043351.1  GI:26335652
KEYWORDS            HTC; CAP trapper.
SOURCE              Mus musculus (house mouse)
  ORGANISM          Mus musculus
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE           1
  AUTHORS           Carninci,P. and Hayashizaki,Y.
  TITLE             High-efficiency full-length cDNA cloning
  JOURNAL           Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE           99279253
   PUBMED           10349636
REFERENCE           2
  AUTHORS           Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
                    Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE             Normalization and subtraction of cap-trapper-selected cDNAs to
                    prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL           Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE           20499374
   PUBMED           11042159
REFERENCE           3
  AUTHORS           Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
                    Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
                    Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
                    Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
                    Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
                    Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
                    Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE             RIKEN integrated sequence analysis (RISA) system--384-format
                    sequencing pipeline with 384 multicapillary sequencer
  JOURNAL           Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE           20530913
   PUBMED           11076861
REFERENCE           4
  AUTHORS           Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
                    Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
                    Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
                    Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
                    Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
                    Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
                    Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
                    Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
                    Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
                    Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
                    Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
                    Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
                    Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
                    Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
                    Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
                    Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
                    Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
                    Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
                    and Hayashizaki,Y.
  TITLE             Functional annotation of a full-length mouse cDNA collection
```

    JOURNAL   Nature 409 (6821), 685-690 (2001)
    MEDLINE   21085660
     PUBMED   11217851
  REFERENCE   5
    AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
              Group Phase I & II Team.
      TITLE   Analysis of the mouse transcriptome based on functional annotation
              of 60,770 full-length cDNAs
    JOURNAL   Nature 420, 563-573 (2002)
  REFERENCE   6 (bases 1 to 2332)
    AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
              Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
              Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
              Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
              Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
              Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
              Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
              Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
              Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
              Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
              Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
              Muramatsu,M. and Hayashizaki,Y.
      TITLE   Direct Submission
    JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
              Physical and Chemical Research (RIKEN), Laboratory for Genome
              Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
              RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
              Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
              URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
              Fax:81-45-503-9216)
    COMMENT   cDNA library was prepared and sequenced in Mouse Genome
              Encyclopedia Project of Genome Exploration Research Group in Riken
              Genomic Sciences Center and Genome Science Laboratory in RIKEN.
              Division of Experimental Animal Research in Riken contributed to
              prepare mouse tissues.
              Please visit our web site for further details.
              URL:http://genome.gsc.riken.go.jp/
              URL:http://fantom.gsc.riken.go.jp/.
   FEATURES   Location/Qualifiers
      source   1..2332
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="FANTOM_DB:A730087N02"
                        /db_xref="taxon:10090"
                        /clone="A730087N02"
                        /tissue_type="cerebellum"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="7 days neonate"
      CDS   379..1809
                        /note="unnamed protein product; hypothetical protein
                        (evidence: rsCDS,ProCrest,decoder,Longest-ORF)
                        putative"
                        /codon_start=1
                        /protein_id="BAC31527.1"
                        /db_xref="GI:26335653"
                        /translation="MALKDTGSGGSTILPISEMVSASSSPGAPLAAAPGPCAPSPFPE

```
                VVELNVGGQVYVTKHSTLLSVPDSTLASMFSPSSPRGGARRRGDLPRDSRARFFIDRD
                GFLFRYVLDYLRDKQLALPEHFPEKERLLREAEFFQLTDLVKLLSPKVTKQNSLNDEC
                CQSDLEDNVSQGSSDALLLRGAAAGAPSGSGAHGVSGVVGGGSAPDKRSGFLTLGYRG
                SYTTVRDNQADAKFRRVARIMVCGRIALAKEVFGDTLNESRDPDRQPEKYTSRFYLKF
                TYLEQAFDRLSEAGFHMVACNSSGTAAFVNQYRDDKIWSSYTEYIFFRPPQKIVSPKQ
                EHEDRKRDKVTDKGSESGTSCNELSTSSCDSHSEASTPQDNPANTQQAAAHQPNTLTL
                DRPSRKAPVQWMPPPDKRRKSELFQSLISKSRETNLSKKKVCEKLSVEEEMKKCIQDF
                KKIHIPDCFPERKRQWQSELLQKYGL"
BASE COUNT        556 a      647 c      620 g      509 t
ORIGIN
```

```
  Query Match                 9.5%;   Score 329;   DB 11;   Length 2332;
  Best Local Similarity      58.5%;   Pred. No. 1.3e-26;
  Matches  781;  Conservative    0;  Mismatches  455;  Indels   99;  Gaps    8;


Qy      560 GGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTT 619
            ||| ||    |   ||     ||||||| |||||| || ||||||||||| || || || |||
Db      481 GGGCCCTGCGCCCCGTCGCCCTTCCCCGAGGTAGTGGAGCTGAATGTTGGCGGCCAGGTT 540

Qy      620 TATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTT 679
            ||| | ||    ||||| || ||  | ||| ||||  |    |||    | |||||
Db      541 TATGTGACCAAGCATTCGACGTTACTCAGCGTCCCGGACAGCACTCTGGCCAGCATGTTC 600

Qy      680 TCCCCAAAGAGAGACACGGCTAATGATC---------------TAGCCAAGGACTCCAAG 724
            || ||    || | ||     | |              | ||| ||||  |
Db      601 TCACCCTCTAGTCCCCGGGGCGGCGCTAGGCGCCGGGGCGACTTGCCCAGGGACAGCCGC 660

Qy      725 GGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGG 784
            |   | || ||||| ||| | || || ||| | || | || | ||||| || || |
Db      661 GCGCGCTTCTTCATCGACCGCGACGGCTTCCTCTTTAGGTACGTGCTGGATTACCTGCGC 720

Qy      785 GACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCT 844
            |||| |||| |||  ||||| || |||||||| || || |  || ||         | |||||
Db      721 GACAAGCAGCTGGCACTGCCCGAGCACTTTCCCGAGAAGGAGAGGCTCCTGCGCGAAGCA 780

Qy      845 GAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGC 904
            || | ||| |||||| | ||| ||||||| || | ||| | | | | |||||| | |
Db      781 GAGTTCTTTCAGCTCACCGACCTGGTCAAGCTGCTGTCGCCCAAGGTCACCAAGCAGAAC 840

Qy      905 CC------AGATGAATTCTGCCACAGTGACTTTGAAGA--------------------- 936
            |         ||||| | |||||| || |||| | || ||
Db      841 TCGCTCAACGATGAGTGCTGCCAGAGCGACCTGGAGGACAACGTTTCCCAGGGCAGCAGC 900

Qy      937 -------------------TGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCT 977
                                | | |   |   | ||| |    | ||||  | | | |
Db      901 GACGCACTGCTGCTGCGTGGGGCGGCGGCTGGCGCGCCCTCGGGCTCTGGGGCACATGGT 960

Qy      978 CCCTG-------------------CTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTG 1018
            | |                     | || | |||         |  || ||| | || ||
Db      961 GTCAGTGGTGTAGTCGGTGGTGGCAGCGCTCCGGACAAGCGCTCTGGGTTCCTCACCCTG 1020

Qy     1019 GGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGA 1078
            || ||| | || || | ||||  ||    ||||       ||||||||||||||||| || |
Db     1021 GGCTACCGTGGCTCTTACACCACGGTGCGAGATAACCAGGCAGATGCCAAGTTCAGGCGT 1080

Qy     1079 GTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACT 1138
            ||  |  | || || |||| ||||| ||||| | || |||||||| || || ||||||||| || |||
```

```
Db    1081 GTGGCGCGCATCATGGTGTGTGGGCGCATAGCCTTGGCCAAGGAGGTCTTTGGGGACACT 1140

Qy    1139 TTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTC 1198
            | |||||| ||  | |||||||| ||    || || |  ||||| |||  | || || |||
Db    1141 CTTAATGAGAGTCGCGACCCTGACCGTCAGCCTGAGAAGTACACATCCCGCTTCTACCTC 1200

Qy    1199 AAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTG 1258
            || |||| || |||| ||| || ||| |||| |||  || ||  || |||||||||||||
Db    1201 AAGTTCACCTACTTGGAGCAGGCGTTCGATCGACTGTCTGAGGCCGGCTTCCACATGGTG 1260

Qy    1259 GCCTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGG 1318
            || || |||||-| || |  || || | ||  ||||||| ||    ||| |||||||||||
Db    1261 GCGTGCAACTCCTCTGGCACTGCCGCCTTTGTCAACCAGTACCGAGACGACAAGATCTGG 1320

Qy    1319 TCAAGCTACACTGAATATGTCTTCTACCGTGAGCCT---TCCAGATGGTCACCCTCACAC 1375
            || |||||||||| |||| || ||| |||      |||   | | ||||||| ||   ||
Db    1321 AGCAGTTACACTGAATACATCTTCTTCCGACCACCTCAGAAAATAGTGTCACCCAAGCAA 1380

Qy    1376 TGCGATTGCTGCTGCAAGAATGGCAAAG---GTGACAAAGAAGGGGAGAGCGGCACGTCT 1432
             ||     |  | ||  |  |||||  ||||||| | | ||||| || || || || ||
Db    1381 GAACATGAAGACAGGAAACGCGACAAAGTCACAGACAAAGGAAGTGAGAGTGGGACTTCC 1440

Qy    1433 TGCAATGACCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAG 1492
            |||||||| |||||||||| |||||| ||||||||| || |||||||||| |||| |||||
Db    1441 TGCAATGAGCTCTCCACATCCAGCTGTGACAGCCACTCAGAGGCCAGCACTCCACAGGAC 1500

Qy    1493 A------CGGTCATCTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGT 1546
            |       |  ||          | |   | | |||| | |||| |   || |||| |
Db    1501 AACCCAGCCAACACTCAGCAGGCTGCAGCTCACCAGCCTAACACCTTAACCTTGGATAGA 1560

Qy    1547 CCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGAC 1606
            ||| || ||| |   ||||| ||| ||| || |  || ||||  | ||| | ||||| ||
Db    1561 CCCTCCAGGAAAGCACCTGTTCAGTGGATGCCACCACCAGACAAGCGCAGAAAAAGTGAA 1620

Qy    1607 TTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCT 1666
             |  | | |    || |    ||| | | ||| | || |    |         | |||||
Db    1621 CTCTTTCAGTCACTCATCAGCAAGTCCCGAGAAACAAATCT------CTCCAAAAAGAAG 1674

Qy    1667 GTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAA 1726
            ||     || ||||| |     | || || || || | || ||||| |||||||    ||||
Db    1675 GTCTGTGAGAAGCTAAGTGTAGAAGAAGAAATGAAAAGTGTATTCAGGATTTTAAAAAA 1734

Qy    1727 AAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGG 1786
            |   | |||||||| | |||||| ||| | | |||| |   |||||||||||||||| |
Db    1735 ATCCATATTCCAGATTGTTTTCCAGAGCGCAAACGCCAGTGGCAATCTGAACTCCTCCAA 1794

Qy    1787 AAGTATCATCTATAA 1801
            || |||    | |||
Db    1795 AAATATGGGTTGTAA 1809
```

RESULT 9
AK047519
LOCUS   AK047519   2343 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
               enriched library, clone:B930082J01 product:hypothetical protein,
               full insert sequence.

```
ACCESSION   AK047519
VERSION     AK047519.1  GI:26092232
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
```

```
  PUBMED     11217851
REFERENCE    5
  AUTHORS    The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
  TITLE      Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
REFERENCE    6  (bases 1 to 2343)
  AUTHORS    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
  COMMENT    cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2343
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B930082J01"
                     /db_xref="taxon:10090"
                     /clone="B930082J01"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     misc_feature    1. .2343
                     /note="hypothetical protein (evidence:
                     rsCDS,ProCrest,decoder,Longest-ORF)"
BASE COUNT      554 a      660 c     624 g      505 t
ORIGIN
```

```
  Query Match                 9.4%;  Score 327.4;  DB 11;  Length 2343;
  Best Local Similarity     58.4%;  Pred. No. 2e-26;
  Matches  780;  Conservative     0;  Mismatches  456;  Indels   99;  Gaps      8;
```

```
Qy    560 GGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTT 619
          ||| ||   |   ||     ||||||| |||||  || ||||||||||||| || || || |||
Db    512 GGGCCCTGCGCCCCGTCGCCCTTCCCCGAGGTAGTGGAGCTGAATGTTGGCGGCCAGGTT 571

Qy    620 TATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTT 679
          ||| | ||     ||||| || ||   | ||| ||||  |   |  |||   |  |||||
Db    572 TATGTGACCAAGCATTCGACGTTACTCAGCGTCCCGGACAGCACTCTGGCCAGCATGTTC 631

Qy    680 TCCCCAAAGAGAGACACGGCTAATGATC--------------TAGCCAAGGACTCCAAG 724
          || ||    ||   ||   | ||     ||            |   |||  ||||  |
Db    632 TCACCCTCTAGTCCCCGGGGCGGCGCTAGGCGCCGGGGCGACTTGCCCAGGGACAGCCGC 691

Qy    725 GGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGG 784
          |   | || |||||| ||| | || || ||| | || | || | |||||| || || |
Db    692 GCGCGCTTCTTCATCGACCGCGACGGCTTCCTCTTTAGGTACGTGCTGGATTACCTGCGC 751

Qy    785 GACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCT 844
          |||| |||| |||  ||||| || ||||||| |||| || | || || |     | |||||
Db    752 GACAAGCAGCTGGCACTGCCCGAGCACTTTCCCGAGAAGGAGAGGCTCCTGCGCGAAGCA 811

Qy    845 GAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGC 904
          || | ||| ||||||  | ||| ||||||| ||| | ||| | ||| | | | ||||||| ||
Db    812 GAGTTCTTTCAGCTCACCGACCTGGTCAAGCTGCTGTCGCCCAAGGTCACCAAGCAGAAC 871

Qy    905 CC------AGATGAATTCTGCCACAGTGACTTTGAAGA-------------------- 936
          |       ||||| | |||||| || ||| | || ||
Db    872 TCGCTCAACGATGAGTGCTGCCAGAGCGACCTGGAGGACAACGTTTCCCAGGGCAGCAGC 931

Qy    937 --------------------TGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCT 977
                              | |  |   |   ||| |    | ||||   | |  |
Db    932 GACGCACTGCTGCTGCGTGGGGCGGCGGCTGGCGCGCCCTCGGGCTCTGGGGCACATGGT 991

Qy    978 CCCTG------------------CTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTG 1018
          | |                  |  || | |||    |  || ||| | || || ||
Db    992 GTCAGTGGTGTAGTCGGTGGTGGCAGCGCTCCGGACAAGCGCTCTGGGTTCCTCACCCTG 1051

Qy   1019 GGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGA 1078
          || ||| | || || | | |||| ||  ||   ||||  |||||||||||||||||| || |
Db   1052 GGCTACCGTGGCTCTTACACCACGGTGCGAGATAACCAGGCAGATGCCAAGTTCAGGCGT 1111

Qy   1079 GTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACT 1138
          || | || || |||| ||||| |   | ||   ||||||| || || ||||||||| || |||
Db   1112 GTGGCGCGCATCATGGTGTGTGGGCGCATAGCCTTGGCCAAGGAGGTCTTTGGGGACACT 1171

Qy   1139 TTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTC 1198
          | ||||| ||  | |||||||||  |  || || ||  ||||| ||| | || ||| |||
Db   1172 CTTAATGAGAGTCGCGACCCTGACCGTCAGCCTGAGAAGTACACATCCCGCTTCTACCTC 1231

Qy   1199 AAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTG 1258
          || |||| |  ||  |||| | ||| || || || ||    |||| |||   || |||||||||||||||
Db   1232 AAGTTCACCTACTTGGAGCAGGCGTTCGATCGACTGTCTGAGGCCGGCTTCCACATGGTG 1291

Qy   1259 GCCTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGG 1318
          || || |||||  || |  || ||  | || ||||||| ||   ||| |||||||||||||||||
Db   1292 GCGTGCAACTCCTCTGGCACTGCCGCCTTTGTCAACCAGTACCGAGACGACAAGATCTGG 1351

Qy   1319 TCAAGCTACACTGAATATGTCTTCTACCGTGAGCCT---TCCAGATGGTCACCCTCACAC 1375
```

```
                    || ||||||||||| |||||| |||      |||        | | |||||||  ||
Db     1352 AGCAGTTACACTGAATACATCTTCTTCCGACCACCTCAGAAAATAGTGTCACCCAAGCAA 1411

Qy     1376 TGCGATTGCTGCTGCAAGAATGGCAAAG---GTGACAAAGAAGGGGAGAGCGGCACGTCT 1432
                 ||     | | ||    | |||||      ||||||| | | ||||| || || ||
Db     1412 GAACATGAAGACAGGAAACGCGACAAAGTCACAGACAAAGGAAGTGAGAGTGGGACTTCC 1471

Qy     1433 TGCAATGACCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAG 1492
            ||||||||| ||||||||||| ||||| ||||||||| || ||||||||| |||| |||||
Db     1472 TGCAATGAGCTCTCCACATCCAGCTGTGACAGCCACTCAGAGGCCAGCACTCCACAGGAC 1531

Qy     1493 A------CGGTCATCTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGT 1546
            |       |    ||       | |   | | |||| | |||| |   || |||| |
Db     1532 AACCCAGCCAACACTCAGCAGGCTGCAGCTCACCAGCCTAACACCTTAACCTTGGATAGA 1591

Qy     1547 CCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGAC 1606
            |||   || ||| |   ||||| |||   ||| | ||    |||| | ||| | || || ||
Db     1592 CCCTCCAGGAAAGCACCTGTTCAGTGGATGCCACCACCAGACAAGCGCAGAAACAGTGAA 1651

Qy     1607 TTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCT 1666
              | | | |    || |        ||| | ||| | || |        |     |||||
Db     1652 CTCTTTCAGTCACTCATCAGCAAGTCCCGAGAAACAAATCT------CTCCAAAAAGAAG 1705

Qy     1667 GTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAA 1726
            ||      || ||||| |    | || || || || | || ||||| |||||||    ||||
Db     1706 GTCTGTGAGAAGCTAAGTGTAGAAGAAGAAATGAAAAGTGTATTCAGGATTTTAAAAAA 1765

Qy     1727 AAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGG 1786
            |   | |||||||||| | ||||| ||| | |||| | ||||||||||||||||| |
Db     1766 ATCCATATTCCAGATTGTTTTCCAGAGCGCAAACGCCAGTGGCAATCTGAACTCCTCCAA 1825

Qy     1787 AAGTATCATCTATAA 1801
            || |||    | |||
Db     1826 AAATATGGGTTGTAA 1840
```

RESULT 10
AK045439
LOCUS   AK045439   2584 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION Mus musculus adult male corpora quadrigemina cDNA, RIKEN
            full-length enriched library, clone:B230119K12 product:hypothetical
            protein, full insert sequence.
ACCESSION   AK045439
VERSION   AK045439.1 GI:26337364
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
```

```
           Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE    Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL  Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE  20499374
   PUBMED  11042159
REFERENCE  3
  AUTHORS  Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
           Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
           Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
           Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
           Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
           Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
           Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE    RIKEN integrated sequence analysis (RISA) system--384-format
           sequencing pipeline with 384 multicapillary sequencer
  JOURNAL  Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE  20530913
   PUBMED  11076861
REFERENCE  4
  AUTHORS  Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
           Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
           Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
           Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
           Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
           Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
           Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
           Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
           Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
           Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
           Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
           Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
           Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
           Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
           Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
           Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
           Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
           Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
           and Hayashizaki,Y.
  TITLE    Functional annotation of a full-length mouse cDNA collection
  JOURNAL  Nature 409 (6821), 685-690 (2001)
  MEDLINE  21085660
   PUBMED  11217851
REFERENCE  5
  AUTHORS  The FANTOM Consortium and the RIKEN Genome Exploration Research
           Group Phase I & II Team.
  TITLE    Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
  JOURNAL  Nature 420, 563-573 (2002)
REFERENCE  6  (bases 1 to 2584)
  AUTHORS  Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
           Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
           Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
           Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
           Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
           Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
           Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
```

Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1. .2584
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B230119K12"
                     /db_xref="taxon:10090"
                     /clone="B230119K12"
                     /sex="male"
                     /tissue_type="corpora quadrigemina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS   262. .1692
                     /note="unnamed protein product; hypothetical protein
                     (evidence: rsCDS,ProCrest,decoder,Longest-ORF)
                     putative"
                     /codon_start=1
                     /protein_id="BAC32368.1"
                     /db_xref="GI:26337365"
                     /translation="MALKDTGSGGSTILPISEMVSASSSPGAPLAAAPGPCAPSPFPE
                     VVELNVGGQVYVTKHSTLLSVPDSTLASMFSPSSPRGGARRRGDLPRDSRARFFIDRD
                     GFLFRYVLDYLRDKQLALPEHFPEKERLLREAEFFQLTDLVKLLSPKVTKQNSLNDEC
                     CQSDLEDNVSQGSSDALLLRGAAAGAPSGSGAHGVSGVVGGGSAPYKRSGFLTLGYRG
                     SYTTVRDNQADAKFRRVARIMVCGRIALAKEVFGDTLNESRDPDRQPEKYTSRFYLKF
                     TYLEQAFDRLSEAGFHMVACNSSGTAAFVNQYRDDKIWSSYTEYIFFRPPQKIVSPKQ
                     EHEDRKRDKVTDKGSESGTSCNELYTSSCDSHSEASTPQDNPANTQQAAAHQPNTLTL
                     DRPSRKAPVQWMPPPDKRRNSELFQSLISKSRETNLSKKKVCEKLSVEEEMKKCIQDF
                     KKIHIPDCFPERKRQWQSELLQKYGL"
BASE COUNT   681 a   636 c   646 g   621 t
ORIGIN

  Query Match   9.3%;   Score 324.2; DB 11;   Length 2584;
  Best Local Similarity   58.3%;   Pred. No. 4.1e-26;
  Matches 778; Conservative   0; Mismatches 458; Indels   99; Gaps   8;

Qy   560 GGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTT 619

```
                  ||| ||   |   ||       |||||||| ||||| || ||||||||||| || || || |||
Db    364 GGGCCCTGCGCCCCGTCGCCCTTCCCCGAGGTAGTGGAGCTGAATGTTGGCGGCCAGGTT 423

Qy    620 TATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTT 679
                  ||| | ||    ||||| || ||  | ||| ||||  |   |    ||| |  |||||
Db    424 TATGTGACCAAGCATTCGACGTTACTCAGCGTCCCGGACAGCACTCTGGCCAGCATGTTC 483

Qy    680 TCCCCAAAGAGAGACACGGCTAATGATC--------------TAGCCAAGGACTCCAAG 724
                  || ||    ||  |  ||    ||          |  ||| ||||  |
Db    484 TCACCCTCTAGTCCCCGGGGCGGCGCTAGGCGCCGGGGCGACTTGCCCAGGGACAGCCGC 543

Qy    725 GGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTCAGG 784
                  |  | || ||||| ||| | || ||  ||| | ||  | ||  | |||||| || || |
Db    544 GCGCGCTTCTTCATCGACCGCGACGGCTTCCTCTTTAGGTACGTGCTGGATTACCTGCGC 603

Qy    785 GACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAAGCT 844
                  |||| |||| |||    ||||| || ||||||| ||  || |   || ||      | |||||
Db    604 GACAAGCAGCTGGCACTGCCCGAGCACTTTCCCGAGAAGGAGAGGCTCCTGCGCGAAGCA 663

Qy    845 GAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAAAGC 904
                  || | ||| |||||| | ||| ||||||| || ||| | ||| | | | |||||| || |
Db    664 GAGTTCTTTCAGCTCACCGACCTGGTCAAGCTGCTGTCGCCCAAGGTCACCAAGCAGAAC 723

Qy    905 CC------AGATGAATTCTGCCACAGTGACTTTGAAGA--------------------- 936
                  |       ||||| | ||||||| || ||| | || ||
Db    724 TCGCTCAACGATGAGTGCTGCCAGAGCGACCTGGAGGACAACGTTTCCCAGGGCAGCAGC 783

Qy    937 -------------------TGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTTCCT 977
                     | | |   |   ||| |   |  | |||| | | | |
Db    784 GACGCACTGCTGCTGCGTGGGGCGGCGGCTGGCGCGCCCTCGGGCTCTGGGGCACATGGT 843

Qy    978 CCCTG------------------CTCCCTGCCGACCGCAAGTGGGGTTTCATTACTGTG 1018
                  | ||                  | ||| | ||    | || ||| | || ||
Db    844 GTCAGTGGTGTAGTCGGTGGTGGCAGCGCTCCGTACAAGCGCTCTGGGTTCCTCACCCTG 903

Qy   1019 GGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGGAGA 1078
                  || ||| | || || | |||| || |||| ||||||||||||||||| || |
Db    904 GGCTACCGTGGCTCTTACACCACGGTGCGAGATAACCAGGCAGATGCCAAGTTCAGGCGT 963

Qy   1079 GTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAAACT 1138
                  || | || || |||| ||||| | ||  |||||||| || ||| |||||||| || |||
Db    964 GTGGCGCGCATCATGGTGTGTGGGCGCATAGCCTTGGCCAAGGAGGTCTTTGGGGACACT 1023

Qy   1139 TTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTATCTC 1198
                  | |||||| ||  | | |||||||| ||   || || | ||||| ||| | || ||| |||
Db   1024 CTTAATGAGAGTCGCGACCCTGACCGTCAGCCTGAGAAGTACACATCCCGCTTCTACCTC 1083

Qy   1199 AAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATGGTG 1258
                  || ||||   || ||||    ||| || |||    |||| ||| || |||||||||||||
Db   1084 AAGTTCACCTACTTGGAGCAGGCGTTCGATCGACTGTCTGAGGCCGGCTTCCACATGGTG 1143

Qy   1259 GCCTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATCTGG 1318
                  || || |||||| || |  || || | || |||||||| ||    ||| ||||||||||||
Db   1144 GCGTGCAACTCCTCTGGCACTGCCGCCTTTGTCAACCAGTACCGAGACGACAAGATCTGG 1203

Qy   1319 TCAAGCTACACTGAATATGTCTTCTACCGTGAGCCT---TCCAGATGGTCACCCTCACAC 1375
                  ||  ||||||||||||  |||||| |||     |||        | |  |||||||   ||
```

```
Db      1204 AGCAGTTACACTGAATACATCTTCTTCCGACCACCTCAGAAAATAGTGTCACCCAAGCAA 1263

Qy      1376 TGCGATTGCTGCTGCAAGAATGGCAAAG---GTGACAAAGAAGGGGAGAGCGGCACGTCT 1432
             ||     | | ||    | |||||    |||||| | | ||||| || || ||
Db      1264 GAACATGAAGACAGGAAACGCGACAAAGTCACAGACAAAGGAAGTGAGAGTGGGACTTCC 1323

Qy      1433 TGCAATGACCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAGGAG 1492
             |||||||| |||| |||||| ||||| ||||||||| || |||||||||| |||| |||||
Db      1324 TGCAATGAGCTCTACACATCCAGCTGTGACAGCCACTCAGAGGCCAGCACTCCACAGGAC 1383

Qy      1493 A------CGGTCATCTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGACCGT 1546
             |       |  ||         | |   | | |||| | |||| |   || ||||   |
Db      1384 AACCCAGCCAACACTCAGCAGGCTGCAGCTCACCAGCCTAACACCTTAACCTTGGATAGA 1443

Qy      1547 CCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGCGAC 1606
             |||  || ||| |  |||||  ||  |||  | |    |||| | ||| | || || ||
Db      1444 CCCTCCAGGAAAGCACCTGTTCAGTGGATGCCACCACCAGACAAGCGCAGAAACAGTGAA 1503

Qy      1607 TTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAAGCT 1666
              |  | | |    || |   ||| | | ||| | ||  |        |     |||||
Db      1504 CTCTTTCAGTCACTCATCAGCAAGTCCCGAGAAACAAATCT------CTCCAAAAGAAG 1557

Qy      1667 GTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTAAAA 1726
             ||    || ||||||     | || || || || | || ||||| |||||||||    ||||
Db      1558 GTCTGTGAGAAGCTAAGTGTAGAAGAAGAAATGAAAAAGTGTATTCAGGATTTTAAAAAA 1617

Qy      1727 AAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGG 1786
             |   | ||||||||| | ||||| ||| | |||| | ||||||||||||||||| |
Db      1618 ATCCATATTCCAGATTGTTTTCCAGAGCGCAAACGCCAGTGGCAATCTGAACTCCTCCAA 1677

Qy      1787 AAGTATCATCTATAA 1801
             || |||    | |||
Db      1678 AAATATGGGTTGTAA 1692
```

RESULT 11
AK042569
LOCUS   AK042569   2555 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:A730006K23 product:hypothetical protein,
            full insert sequence.
ACCESSION   AK042569
VERSION   AK042569.1 GI:26335190
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 2555)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,

           Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
           Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
           Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
  COMMENT   cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site for further details.
           URL:http://genome.gsc.riken.go.jp/
           URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1. .2555
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A730006K23"
                     /db_xref="taxon:10090"
                     /clone="A730006K23"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="7 days neonate"
     CDS   262. .1695
                     /note="unnamed protein product; hypothetical protein
                     (evidence: rsCDS,ProCrest,decoder,Longest-ORF)
                     putative"
                     /codon_start=1
                     /protein_id="BAC31296.1"
                     /db_xref="GI:26335191"
                     /translation="MALKDTGSGGSTILPISEMVSASSSPGAPLAAAPGPCAPSPFPE
                     IVELNVGGQVYVTKHSTLLSVPDSTLASMFSPSSPRGRAPRRRGDLPRDSRARFFIDR
                     DGFLFRYVLDYLRDKQLALPEHFPEKERLLREAEFFQLTDLVKLLSPKVTKQNSLNDE
                     CCQSDLEDNVSQGSSDALLLRGAAAGAPSGSGAHGVSGVVGGGSAPDKRSGFLTLGYR
                     GSYTTVRDNQADAKFRRVARIMVCGRIALAKEVFGDTLNESRDPDRQPEKYTSRFYLK
                     FTYLEQAFDRLSEAGFHMVACNSSGTAAFVNQYRDDKIWSSYTEYIFFRPPQKIVSPK
                     QEHEDRKRDKVTDKGSESGTSCNELSTSSCDSHSEASTPQDNPANTQQAAAHQPNTLT
                     LDRPSRKAPVQWMPPPDKRRNSELFQSLISKSRETNLSKKKVCEKLSVEEEMKKCIQD
                     FKKIHIPDCFPERKRQWQSELLQKYGL"
BASE COUNT   666 a   638 c   643 g   608 t
ORIGIN

     Query Match   9.3%;   Score 322.8;   DB 11;   Length 2555;
     Best Local Similarity   58.2%;   Pred. No. 5.8e-26;
     Matches   779; Conservative   0; Mismatches   457; Indels 102; Gaps   8;

Qy   560 GGGTCCGCAGTTCCCAACTCCTTCCCTGAGGTGGTAGAGCTGAATGTCGGGGGTCAAGTT 619
             ||| || | || ||||||| ||| | || ||||||||||| || || || |||
Db   364 GGGCCCTGCGCCCCGTCGCCCTTCCCCGAGATAGTGGAGCTGAATGTTGGCGGCCAGGTT 423

```
Qy    620 TATTTTACTCGCCATTCCACATTGATAAGCATCCCTCATTCCCTCCTGTGGAAAATGTTT 679
          ||| | ||      |||||| || ||  | ||| ||||  |     |    |||   | |||||
Db    424 TATGTGACCAAGCATTCGACGTTACTCAGCGTCCCGGACAGCACTCTGGCCAGCATGTTC 483

Qy    680 TCCCCAAAGAGAGACACGGCTAATGATC------------------TAGCCAAGGACTCC 721
          || ||      ||    | ||      | |             |   | ||| ||||  |
Db    484 TCACCCTCTAGTCCCCGGGGGCGGGCGCCTAGGCGCCGGGGCGACTTGCCCAGGGACAGC 543

Qy    722 AAGGGAAGGTTTTTCATTGACAGAGATGGATTCTTGTTCCGTTATATTCTGGACTATCTC 781
          |    | || ||||| ||| | || || ||| | || | || | ||||| || ||
Db    544 CGCGCGCGCTTCTTCATCGACCGCGACGGCTTCCTCTTTAGGTACGTGCTGGATTACCTG 603

Qy    782 AGGGACAGGCAGGTGGTCCTGCCTGATCACTTTCCAGAAAAAGGAAGACTGAAAAGGGAA 841
          | |||| |||| |||  ||||| || |||||||| || || | || ||     | |||
Db    604 CGCGACAAGCAGCTGGCACTGCCCGAGCACTTTCCCGAGAAGGAGAGGCTCCTGCGCGAA 663

Qy    842 GCTGAATACTTCCAGCTCCCAGACTTGGTCAAACTCCTGACCCCCGATGAAATCAAGCAA 901
          || || | ||| |||||| | |||| |||||||| || ||| | ||| | | | ||||||
Db    664 GCAGAGTTCTTTCAGCTCACCGACCTGGTCAAGCTGCTGTCGCCCAAGGTCACCAAGCAG 723

Qy    902 AGCCC------AGATGAATTCTGCCACAGTGACTTTGAAGA------------------ 936
          | | |       ||||| | |||||| || ||| | || ||
Db    724 AACTCGCTCAACGATGAGTGCTGCCAGAGCGACCTGGAGGACAACGTTTCCCAGGGCAGC 783

Qy    937 --------------------TGCCTCCCAAGGAAGCGACACAAGAATCTGCCCCCCTT 974
                              | | |   |     | |||| |     | ||||   | | |
Db    784 AGCGACGCACTGCTGCTGCGTGGGGCGGCGGCTGGCGCGCCCTCGGGCTCTGGGGCACAT 843

Qy    975 CCTCCCTG-------------------CTCCCTGCCGACCGCAAGTGGGGTTTCATTACT 1015
          | | |                      | || ||| |     | || ||| | ||
Db    844 GGTGTCAGTGGTGTAGTCGGTGGTGGCAGCGCTCCGGACAAGCGCTCTGGGTTCCTCACC 903

Qy   1016 GTGGGTTACAGAGGATCCTGCACCTTGGGCAGAGAGGGACAGGCAGATGCCAAGTTTCGG 1075
          |||| ||| | || || | ||||  || ||||  | ||||||||||||||||||| ||
Db    904 CTGGGCTACCGTGGCTCTTACACCACGGTGCGAGATAACCAGGCAGATGCCAAGTTCAGG 963

Qy   1076 AGAGTTCCCCGGATTTTGGTTTGTGGAAGGATTTCCTTGGCAAAAGAAGTCTTTGGAGAA 1135
          | ||  |  | || ||  |||| |||||| ||||| | ||| |||||||| || || |||||||| ||
Db    964 CGTGTGGCGCGCATCATGGTGTGTGGGCGCATAGCCTTGGCCAAGGAGGTCTTTGGGGAC 1023

Qy   1136 ACTTTGAATGAAAGCAGAGACCCTGATCGAGCCCCAGAAAGATACACCTCCAGATTTTAT 1195
          ||| | ||||| ||  | ||||||||| ||   || || | ||||| ||| | || ||
Db   1024 ACTCTTAATGAGAGTCGCGACCCTGACCGTCAGCCTGAGAAGTACACATCCCGCTTCTAC 1083

Qy   1196 CTCAAATTCAAGCACCTGGAAAGGGCTTTTGATATGTTGTCAGAGTGTGGATTCCACATG 1255
          ||||| ||||    || ||||   ||| || |||    |||| ||| || |||||||||||
Db   1084 CTCAAGTTCACCTACTTGGAGCAGGCGTTCGATCGACTGTCTGAGGCCGGCTTCCACATG 1143

Qy   1256 GTGGCCTGTAACTCATCGGTGACAGCATCTTTCATCAACCAATATACAGATGACAAGATC 1315
          ||||| || ||||| || |  || ||  | |  |  ||||||| ||     ||| |||||||||
Db   1144 GTGGCGTGCAACTCCTCTGGCACTGCCGCCTTTGTCAACCAGTACCGAGACGACAAGATC 1203

Qy   1316 TGGTCAAGCTACACTGAATATGTCTTCTACCGTGAGCCT---TCCAGATGGTCACCCTCA 1372
          |||   || || |||||||||| || ||||| ||||    |   ||| | |||||||
Db   1204 TGGAGCAGTTACACTGAATACATCTTCTTCCGACCACCTCAGAAAATAGTGTCACCCAAG 1263
```

```
Qy      1373 CACTGCGATTGCTGCTGCAAGAATGGCAAAG---GTGACAAAGAAGGGGAGAGCGGCACG 1429
             ||      ||     | | ||     | |||||      |||||||| | | |||||| || ||
Db      1264 CAAGAACATGAAGACAGGAAACGCGACAAAGTCACAGACAAAGGAAGTGAGAGTGGGACT 1323

Qy      1430 TCTTGCAATGACCTCTCCACATCTAGCTGCGACAGCCAGTCTGAGGCCAGCTCTCCCCAG 1489
             || ||||||||| |||||||||||| ||||| |||||||| || |||||||||| |||| |||
Db      1324 TCCTGCAATGAGCTCTCCACATCCAGCTGTGACAGCCACTCAGAGGCCAGCACTCCACAG 1383

Qy      1490 GAGA------CGGTCATCTGTGGTCCCGTGACACGCCAGACCAACATCCAGACTCTGGAC 1543
             || |       |   ||      || |   | | |||| | |||| |    || ||||
Db      1384 GACAACCCAGCCAACACTCAGCAGGCTGCAGCTCACCAGCCTAACACCTTAACCTTGGAT 1443

Qy      1544 CGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAGAGATGCGGCGGAAAAGC 1603
              |  |||  ||  |||  |  |||||  |||   |||  | | |    |||| |  ||| | | ||
Db      1444 AGACCCTCCAGGAAAGCACCTGTTCAGTGGATGCCACCACCAGACAAGCGCAGAAACAGT 1503

Qy      1604 GACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACATGAGCAGCAAAAAAAAA 1663
             || | | || |   || |         ||| | ||| | || |    |      ||||||
Db      1504 GAACTCTTTCAGTCACTCATCAGCAAGTCCCGAGAAACAAATCT------CTCCAAAAAG 1557

Qy      1664 GCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAATGTATCCAGGATTTCCTA 1723
               ||      || |||||       | || || ||  || | || ||||| ||||||||    |
Db      1558 AAGGTCTGTGAGAAGCTAAGTGTAGAAGAAGAAATGAAAAGTGTATTCAGGATTTTAAA 1617

Qy      1724 AAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTA 1783
             ||||     | |||||||||| | |||||  ||| | ||||   |  ||||||||||||||| |
Db      1618 AAAATCCATATTCCAGATTGTTTTCCAGAGCGCAAACGCCAGTGGCAATCTGAACTCCTC 1677

Qy      1784 AGGAAGTATCATCTATAA 1801
             || |||    | |||
Db      1678 CAAAAATATGGGTTGTAA 1695
```

RESULT 12
BQ713664
LOCUS   BQ713664   973 bp   mRNA   linear   EST 16-JUL-2002
DEFINITION   AGENCOURT_8480138 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6310836
              5', mRNA sequence.
ACCESSION   BQ713664
VERSION   BQ713664.1 GI:21852563
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 973)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Susan L. Sullivan, PhD.
               cDNA Library Preparation: ResGen, Invitrogen Corp
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Agencourt Bioscience Corporation
               Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:

Plate: LLAM13733 row: m column: 13
High quality sequence start: 23
High quality sequence stop: 592.

```
FEATURES              Location/Qualifiers
     source           1. .973
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:6310836"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_129"
                      /note="Organ: olfactory epithelium; Vector:
                      pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                      unidirectionally. Primer: Oligo dT. Average insert size
                      2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                      is a NIH_MGC Library."
BASE COUNT      275 a      216 c      192 g      288 t       2 others
ORIGIN
```

```
  Query Match              8.9%;   Score 310;   DB 13;   Length 973;
  Best Local Similarity   68.8%;   Pred. No. 2.4e-24;
  Matches  681; Conservative    0;  Mismatches  242;  Indels   67;  Gaps   16;
```

```
Qy     1690 GGAGGAGCTGGAGAAATGTATCCAGGATTTCCTA--AAAAAAAAAATTCCAGATCGGTTT 1747
            ||| |||||||||||||||||||||   |  || ||    || | |||||||||||||| ||
Db        1 GGAAGAGCTGGAGAAATGTATCNCAGGATTTCTNTGAAGATAAAAATTCCAGATCGCTTC 60

Qy     1748 CCTGAGAGAAAACATCCTTGGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGG 1807
            ||||||| |||||||||||||||| |||||||||||| ||||||||||||||||| ||| |
Db       61 CCTGAGCGAAAACATCCTTGGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGGAG 120

Qy     1808 GCTGGGGGCGGGGAAAAAAAAAAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAAT 1867
            |   | ||                    |    | || ||||||| ||||||   |||||||
Db      121 GGCTGTGG---------------GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAG 165

Qy     1868 TCATATTTTAAAGGAAAAAAATACAACTAATGATGCACATTTCTTAGAACACAATAGTCC 1927
            ||||| |||||||||||| | ||||||||| | | ||||||| ||||| |||||| ||||
Db      166 ACATATGTTAAAGGAAAAATA-ACAACTAACGGTCCACATTTGTTAGATCACAAT-GTCC 223

Qy     1928 ATTGATATACTACTGCCTACTTTACCTAGTTCACCTTAACATGTAAATCCACAGGGTAGA 1987
            |||||| |||||||||||||||| ||||| |||||||||| |||||||||||||||||||
Db      224 ATTGATGTACTACTGCCTACTTTGCCTAGCTCACCTTAACGTGTAAATCCACAGGGTAGA 283

Qy     1988 TTTCTTTCTAGATGTGGAAGTACAAGAAAATCTTTTTTAGT--TATTTGTTTGTTTACTT 2045
            |||||||||||||||||||| | || |   || | || | |   ||| |||||||||
Db      284 TTTCTTTCTAGATGTGGAACCAGAAACGAGCTCTTAGTTGTCCTTGTCTTTTATTTACTT 343

Qy     2046 CGTCCCATGTGCTAACTATCTT-ATATATAATGAGAGCCAGCTACGTAAAAGTAGCTGAG 2104
            |||||||||||| |  ||||| | ||| || |||  |||||||||| ||     ||||||| |
Db      344 GGTCCCATGTGCTGAGAATCTTAAGATACAACAAGAACCAGCTACGTGTGAGTAGCTCAC 403

Qy     2105 AGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTT--------------------- 2141
            |||| ||||||| ||||| ||||||||||   |||||||||
Db      404 AGGCTTTGGGAATCATTGATCCCAAACCAGGTTTTTTTGTTTTGTTTTGTTTTGTTTTGT 463

Qy     2142 ---CTCTCATCCTTCTACCTCCCTCCTTTGA--ATGAGGGTATGGTAGAAAAAGATCTGG 2196
```

```
                  |||||||  ||||  |||||  || ||||  | ||  | |  || |||  |||| |
Db       464 TTACTCTCATTTTTCTGCCTCCTCCCCTTGACCAAGAATGGACAGTTGAAGGAGATATAA 523

Qy      2197 CCCAATGGCATAAGTTTGGAATTTTTAATTTTGGTTTTTCCT---TTTGTTTATGGGGTT 2253
                  |||  ||||  || |||  ||| || |  ||||   ||| || |   ||||||||||I'IIIII
Db       524 CCCGGTGGCTTATGTTAAGAAATTATCCTTTTCCCTTTCCTTTTGTTTGTTTATGGGGTT 583

Qy      2254 --GGGGGGAATGGCAGATTTATATGACTTTTCACTCAAATCTATATGTGCCAGTTTATAT 2311
                  ||||  ||||||||| ||||  ||||| |||||||||  ||||||| |||||||||| || |||
Db       584 GAGGGGAGAATGGCAAATTTGTATGATTTTTCACTAAAATCTCTATGTGCCAGGTTCTAT 643

Qy      2312 TGACTCCGTATGCATGAGTATTTGTG---CAACACAAGCACAACTAAGTATGTATATACA 2368
                  |||||  |||||||||||  || ||   |||    || | | | |  |||| |||  |||
Db       644 TGACTTTGTATGCATGAGCCGTTCTGACACAAGCACAGTATATGTCTGTATATATGCACA 703

Qy      2369 CATGACGCACACGATGCCAGGGCCTAGACCTCCCAAGGGCTGTGCTCCTGCTCCCAGCAG 2428
                  |  | |||||||||      |||||   |  |  |  |||||||   || | ||  |||| |
Db       704 AAGAATGCACACGACCTAAGGGC--TGGACAGCAGAGGGCTAACATCTTACTATCAGCTG 761

Qy      2429 CCCTCTCTTAGAATATTTCAGATGGATGAGCTTCTGACTCTTTCTTAAAATTCTTTTGGG 2488
                  ||| ||     |||  |||||||||   |  ||| ||||  ||  ||| || |||   ||||
Db       762 CCC-CTACAAGAGCATTTCAGA-CAACAAGCCTCTGGCTATTTATT-AAACCCTCCTGGG 818

Qy      2489 AAGATTTCCCAGCCTTTCTT----CACAACACTTTCTAACATCAAATGACTCTCATCATC 2544
                  | ||||||||||||||  | |        ||||  ||| || | || |    | |||||||
Db       819 CAAATTTCCCAGCCTCCCCTGGCAGGCAACCTTTTTAAAGCTGAATTAGGCCCCATCATC 878

Qy      2545 AACAAATTGTATTCCTTATTGTGAAATTAATACCCTCAGGCTCCATTTTACTGCTTTGCT 2604
                  || ||||||     ||||| |||     ||| || |||| |||||||||| ||  |      |
Db       879 AAAAAATTCCCTTCCTAATTTGAAAAAAAAAACCCCCAGGCTCCCTTGGAAATAATAAGG 938

Qy      2605 CTTTGTCTGCATTAAGAGAGGATGAGGAGA 2634
                  |  | || ||   | |||  || ||| |
Db       939 CCCTTTCCCCACCTAAAGAACCTGGGGAAA 968
```

RESULT 13
AA332022

              Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
              Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
              Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
              , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
              Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
              Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
              ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
              Venter,J.C.
  TITLE   Initial assessment of human gene diversity and expression patterns
              based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
   PUBMED   7566098
  COMMENT   Contact: Kerlavage, AR
              Bioinformatics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850 USA
              Tel: 3018699056
              Fax: 3018699423
              Email: arkerlav@tigr.org
              For clone availability, additional sequence and expression
              information related to this EST, please check the TIGR Human Gene
              Index (http://www.tigr.org/tdb/hgi/hgi.html)
              Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1. .319
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):133650"
                     /db_xref="taxon:9606"
                     /dev_stage="embryo, 8 wks"
                     /clone_lib="Embryo, 8 week I"
                     /note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT   99 a   61 c   47 g   112 t
ORIGIN

   Query Match   8.8%;   Score 305.4;   DB 9;   Length 319;
   Best Local Similarity   99.4%;   Pred. No. 1.4e-23;
   Matches 317; Conservative   0; Mismatches   1; Indels   1; Gaps   1;

Qy   1842 TTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGAT 1901
             |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   1 TTTGAAATTAACCTCCTAAAAGGAATTCATATTTTAAAGGAAAAAAATACAACTAATGAT 60

Qy   1902 GCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCAC 1961
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCACATTTCTTAGAACACAATAGTCCATTGATATACTACTGCCTACTTTACCTAGTTCAC 120

Qy   1962 CTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTT 2021
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTTAACATGTAAATCCACAGGGTAGATTTCTTTCTAGATGTGGAAGTACAAGAAAATCTT 180

Qy   2022 TTTTAGTTATTTGTTTGTTTACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAG 2081
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTTTAGTTATTTGTTTGTTTACTTCGTCCCATGTGCTAACTATCTTATATATAATGAGAG 240

```
Qy    2082 CCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGG-TTTTT 2140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     241 CCAGCTACGTAAAAGTAGCTGAGAGGCCTTGGGAGTCATTTATCCCAAACTGGGTTTTTT 300

Qy    2141 TCTCTCATCCTTCTACCTC 2159
           |||||||||||||||||||
Db     301 TCTCTCATCCTTCTACCTC 319
```

RESULT 14
BY706433
```
LOCUS       BY706433              424 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION  BY706433 RIKEN full-length enriched, adult male testis Mus musculus
            cDNA clone 1700026A08 5', mRNA sequence.
ACCESSION   BY706433
VERSION     BY706433.1  GI:27117598
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
```

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
Submission

  Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
  Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
  RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
  Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
  cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
  Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

```
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1700026A08"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /clone_lib="RIKEN full-length enriched, adult male testis"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites. "
```

BASE COUNT   133 a   98 c   114 g   79 t
ORIGIN

    Query Match   8.3%;   Score 286.2;   DB 14;   Length 424;
    Best Local Similarity   82.0%;   Pred. No. 1.3e-21;
    Matches 360; Conservative   0; Mismatches   63; Indels   16; Gaps   2;

Qy   1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526
             |||| ||||||||||||||| ||||||||||| ||||||||| || || || ||||||| ||
Db   2 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 61

Qy   1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
             |||||||||||||||| || |||||||||| || || || |||||||||||||||||||||||||||
Db   62 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 121

Qy   1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
             ||||| ||||||||||| ||| | |||||||| |||||| |||||||||||||| |||||||
Db   122 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA 181

Qy   1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
             | ||||||||||||| ||||||| || |||||||||||| || ||||| ||||||||||||||
Db   182 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT 241

Qy   1707 GTATCCAGGATTTCCTAAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAAACATCCTT 1766
             |||||||||||||| | || | |||||||||||||| || ||||||| |||||||||||||
Db   242 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAAACATCCTT 301

Qy   1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAAA 1826
             |||| |||||||||||| ||||||||||||||||| ||| || | ||
Db   302 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGGAGGGCTGTGG---------- 350

Qy   1827 AAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
             | | || |||||||| |||||| ||||||| |||||| ||||| |||||||||||||
Db   351 ----GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAGACATATGTTAAAGGAAAAA 406

Qy   1887 AATACAACTAATGATGCAC 1905
             | |||||||| | | |||
Db   407 TA-ACAACTAACGGTCCAC 424

RESULT 15
AK006368
LOCUS   AK006368   422 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus adult male testis cDNA, RIKEN full-length enriched
             library, clone:1700026A08 product:inferred: RIKEN cDNA 4930434H12
             gene / putative [Mus musculus], full insert sequence.
ACCESSION   AK006368
VERSION   AK006368.1 GI:12839431
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)

MEDLINE   99279253
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 422)
AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,

            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
  COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA
            was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
            end: SstI. Host: SOLR.
FEATURES   Location/Qualifiers
     source   1. .422
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:1700026A08"
                     /db_xref="MGI:1893081"
                     /db_xref="taxon:10090"
                     /clone="1700026A08"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature   1. .422
                     /note="inferred: RIKEN cDNA 4930434H12 gene / putative
                     [Mus musculus] (UniGene|Mm.46143, TIGR-MGI1|TC1870,
                     evidence: UG/TGI)"
                     /db_xref="MGI:1914659"
BASE COUNT   132 a   97 c   114 g   79 t
ORIGIN

  Query Match   8.2%;   Score 286; DB 11;   Length 422;
  Best Local Similarity   82.2%;   Pred. No. 1.4e-21;
  Matches 347; Conservative   0; Mismatches   60; Indels   15; Gaps   1;

Qy   1467 AGTCTGAGGCCAGCTCTCCCCAGGAGACGGTCATCTGTGGTCCCGTGACACGCCAGACCA 1526

```
                     ||||  |||||||||||||||  ||||||||||||  |||||||||  ||  ||  ||  ||||||||  ||
Db          2 AGTCAGAGGCCAGCTCTCCGCAGGAGACGGTGATCTGTGGGCCTGTAACGCGCCAGAGCA 61

Qy       1527 ACATCCAGACTCTGGACCGTCCCATCAAGAAGGGCCCTGTCCAGCTGATCCAACAGTCAG 1586
                     |||||||||||||||||  ||  ||||||||||||  ||  ||  ||  |||||||||||||||||||||||
Db         62 ACATCCAGACTCTGGATCGGCCCATCAAGAAAGGTCCGGTGCAGCTGATCCAACAGTCAG 121

Qy       1587 AGATGCGGCGGAAAAGCGACTTACTCCGGATTCTGACTTCAGGCTCCAGGGAATCGAACA 1646
                     |||||  ||||||||||  |||  |  ||||||||  ||||||  ||||||||||||||||  |||||||
Db        122 AGATGAGGCGGAAAAGTGACCTGCTCCGGACTCTGACGTCAGGCTCCAGGGAGTCGAACA 181

Qy       1647 TGAGCAGCAAAAAAAAAGCTGTTAAAGAAAAGCTCTCAATTGAGGAGGAGCTGGAGAAAT 1706
                     |  ||||||||||||  |||||||    ||  |||||||||||  ||  |||||  ||||||||||||
Db        182 TAAGCAGCAAAAAGAAAGCTGCGAAGGAAAAGCTCTCCATCGAGGAAGAGCTGGAGAAAT 241

Qy       1707 GTATCCAGGATTTCCTAAAAAAAAAAAAATTCCAGATCGGTTTCCTGAGAGAAACATCCTT 1766
                     ||||||||||||  |  ||  |  ||||||||||||||  ||  ||||||  ||||||||||||
Db        242 GTATCCAGGATTTCTTGAAGATAAAAATTCCAGATCGCTTCCCTGAGCGAAACATCCTT 301

Qy       1767 GGCAATCTGAACTTTTAAGGAAGTATCATCTATAAGGGAGGGCTGGGGGCGGGGAAAAA 1826
                     ||||  |||||||||||  ||||||||||||||||||  |||  ||  |  ||
Db        302 GGCAGTCTGAACTTTTACGGAAGTATCATCTATAGGGGGAGGGCTGTGG---------- 350

Qy       1827 AAAAAAAGAGTCATTTTGAAATTAACCTCATAAAAGGAATTCATATTTTAAAGGAAAAA 1886
                     |     |  ||  |||||||||  ||||||      |||||||   |||||  |||||||||||
Db        351 ----GTAGTCGCCACTTTGAAATAAACCTCCCCAAAGGAAGACATATGTTAAAGGAAAAA 406

Qy       1887 AA 1888
                     |
Db        407 TA 408
```

Search completed: January 28, 2004, 23:02:16
Job time : 6872 secs